

1999

## Teacher efficacy and behavior: their relationship and impact on student learning

Linda Diane Avery

*College of William & Mary - School of Education*

Follow this and additional works at: <https://scholarworks.wm.edu/etd>



Part of the [Curriculum and Instruction Commons](#), [Educational Psychology Commons](#), and the [Teacher Education and Professional Development Commons](#)

---

### Recommended Citation

Avery, Linda Diane, "Teacher efficacy and behavior: their relationship and impact on student learning" (1999). *Dissertations, Theses, and Masters Projects*. Paper 1539618728.

<https://dx.doi.org/doi:10.25774/w4-jbdp-7t82>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact [scholarworks@wm.edu](mailto:scholarworks@wm.edu).

## **INFORMATION TO USERS**

**This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.**

**The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.**

**In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.**

**Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.**

**Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.**

**Bell & Howell Information and Learning  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
800-521-0600**

**UMI<sup>®</sup>**



**TEACHER EFFICACY AND BEHAVIOR:  
THEIR RELATIONSHIP AND IMPACT ON STUDENT LEARNING**

---

A Dissertation

Presented to

The Faculty of the School of Education

The College of William and Mary

---

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

---

by

Linda D. Avery

December, 1999

**UMI Number: 9961159**

**UMI<sup>®</sup>**

---

**UMI Microform 9961159**

**Copyright 2000 by Bell & Howell Information and Learning Company.**

**All rights reserved. This microform edition is protected against  
unauthorized copying under Title 17, United States Code.**

---

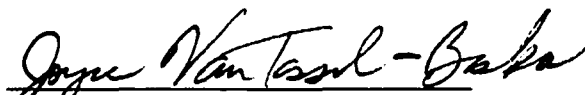
**Bell & Howell Information and Learning Company  
300 North Zeeb Road  
P.O. Box 1346  
Ann Arbor, MI 48106-1346**

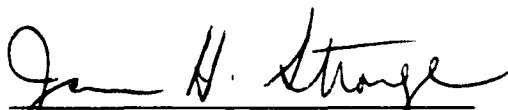
**TEACHER EFFICACY AND BEHAVIOR:  
THEIR RELATIONSHIP AND IMPACT ON STUDENT LEARNING**

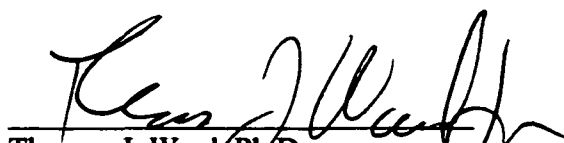
by Linda D. Avery

---

Approved December, 1999

  
Joyce VanTassel-Baska, Ed.D.  
Chair of Dissertation Committee

  
James H. Stronge, Ph.D.

  
Thomas J. Ward, Ph.D.

## TABLE OF CONTENTS

Acknowledgments .....	v
List of Tables .....	viii
List of Figures .....	x
Abstract .....	xi
Half-Title Page .....	xiii
Chapter I: Statement of the Problem .....	1
Introduction .....	1
Conceptual Framework .....	3
Purpose and Research Questions .....	4
Rationale for the Study .....	5
Definition of Terms .....	6
Synopsis of Methodology .....	9
Significance of the Research .....	10
Context for the Research .....	11
Limitations and Delimitations .....	11
Chapter II: Review of the Literature .....	13
Introduction .....	13
Strand I: Teacher Efficacy and Its Relationship to Teacher Behavior and Student Learning .....	16
Strand II: Teacher Change through Staff Development and Transfer of Learning ....	28
Strand III: Curriculum Improvements for High Ability Learners .....	32
Strand IV: Economically Disadvantaged and Culturally Diverse Gifted Students ....	41
Implications from the Overall Review of the Literature .....	46
Chapter III: Methodology .....	51
Conceptual Framework .....	51
Research Questions and Instruments .....	52
Research Design .....	53
Site Selection .....	54
Description of the Intervention .....	56
Sample .....	57
Instrumentation .....	57
Data Collection .....	60
Data Analysis Procedures .....	61
Time Frame for the Study .....	62
Confidentiality and Other Ethical Considerations .....	63

<b>Chapter IV: Findings</b>	<b>64</b>
Introduction	64
Sample	65
Factor Analysis Component of the Study	69
Question #1 Results	75
Question #2 Results	85
Question #3 Results	92
Question #4 Results	99
Summary of Findings	102
<b>Chapter V: Discussion, Conclusions, and Implications</b>	<b>107</b>
Teacher Efficacy and Its Relationship to Teacher Behavior and Student Learning	107
Discussion Related to Curriculum Innovation	115
Conclusion	121
Implications for Research and Practice	122
Summary	125
<b>Appendix: Instruments Used in the Study</b>	<b>128</b>
Demographic Information	129
General Teaching Efficacy Scale	129
Social Studies Teaching Efficacy Scale	130
Classroom Observation Form - External Observer	132
Teacher Self-Report Inventory	135
Concept Learning Test - Grade 2	138
Concept Learning Test - Grade 4	144
Critical Thinking Test - Grade 2	150
Critical Thinking Test - Grade 4	155
Combined Concept Learning and Critical Thinking Test - Grade 7	161
<b>References</b>	<b>169</b>
<b>Vita</b>	<b>183</b>



## ACKNOWLEDGMENTS

There are some lines by the poet Rilke that I believe capture the spirit of my decision in late mid-life to pursue a doctorate:

. . .be patient toward all that is unsolved in your heart and try to love the *questions themselves* like locked rooms and like books that are written in a foreign tongue . . .

It was therefore fitting that the capstone of this educational experience for me was the preparation of a study to explore new questions of interest in an environment that encouraged curiosity, concentration, reflection, and risk-taking. It was an experience that confirmed for me that the questions we raise are as important as the answers we find. Within the context of a community of scholars and learners, I inquired more broadly, probed more deeply, and discerned more acutely.

I would like to express my gratitude to those members of the academic community here at the College of William and Mary who made this journey challenging, stimulating, and worthwhile. Every course I had the privilege to take was taught by a faculty member committed to the importance of the teaching-learning enterprise. The faculty is characterized by the high quality of their pedagogical expertise, their investment in the classroom experience, and their availability beyond their posted office hours. While I cannot acknowledge each of them individually, I can thank Dean Virginia McLaughlin for shaping a culture in the School of Education that puts student learning at the heart of the school's mission.

Special accolades are sent to two of the members of my dissertation committee, Dr. James Stronge and Dr. Thomas Ward. I selected these individuals not only for their subject expertise, with each contributing important knowledge of their specialty areas, but because of

who they are as people. Dr. Stronge is the consummate professional, integrating teaching, scholarship, and service with such dexterity that he makes these competing demands on his time look natural and effortless. He is also an undaunted optimist whose enthusiasm for his job is contagious. Dr. Ward is the resident statistical wizard. He makes quantitative paradigms accessible to the narrative mind and has tremendous patience in answering myriad questions. Although promoted to the Associate Deanship during the execution of this study, he stayed actively involved through the data analysis phase, always helping me to understand the implications of the decisions I was making.

Most of all, I would like to thank Dr. Joyce VanTassel-Baska, the chair of my committee, my advisor, and my mentor. Dr. VanTassel-Baska encouraged me to pursue graduate work at several points in my career and made William and Mary the optimal choice for fulfilling this goal. It is difficult to find the words to express my regard and esteem for this remarkable person. Her leadership in the field of gifted education has impacted the lives of thousands of students and educators, and her accomplishments are prolific. In spite of her extensive publication record, her numerous awards and honors, and her international reputation for excellence, I think the appellation that most pleases her professionally is that of “teacher.” Dr. VanTassel-Baska is the quintessential teacher, one who understands that teaching and learning are reciprocal processes, and that the next time that we do something, we can always do it better. She believes in the power of education to shape human destiny, and she approaches every professional encounter as an opportunity to make a difference. She set high expectations for my performance, provided constructive feedback on my work, created challenges that helped me to grow, and modeled reflective practice. I would not be standing at this threshold of my life without her guidance, her encouragement, her determination, and her wisdom.

My gratitude must also be extended to several of my fellow graduate students and colleagues. Molly Pence assisted me with data collection and entry, Catherine Little with editing, and Beth Stokes with manuscript preparation. Deborah Holland, the Gifted Coordinator for Norfolk Public Schools, and Jane Ann Snyder, her assistant, rallied the Norfolk staff to help with the observation and student testing processes. Like coaches on the sidelines, they inspired me to stay focused on the goal. All of their contributions made this dissertation stronger.

I also want to acknowledge my family for their support both emotional and financial. My siblings and their spouses stayed in constant touch with me during my sojourn in Virginia, listening to my tales of success and failure, advising me to stay the course, offering their homes and their hospitality for my infrequent retreats from academia. I appreciated their concern and their care packages. For Sandy, Mike, Alan, Lorri, Susan, Jennifer, Scott, and Rena, my heartfelt thanks!

My parents, God bless them, have always shown me unconditional love. Before I finalized my decision to seek a doctorate on a full-time basis, my mother called to remind me of the importance of having dreams and the persistence to pursue them. My dad then chimed in with his full support and assured me that they were both behind me. Their words of encouragement will reverberate through the archways of my mind for the rest of my life.

## LIST OF TABLES

1.	Data Collection Framework .....	61
2.	Grade Level by Group .....	65
3.	Gender by Group .....	66
4.	Age by Group .....	66
5.	Level of Education by Group .....	67
6.	Level of Experience by Group .....	68
7.	Statistical Analysis of Four Sample Variables .....	69
8.	Factor Loadings, Means, and Standard Deviations for the Teacher Efficacy Scale–Short Form .....	70
9.	Factor Loadings, Means, and Standard Deviations for Social Studies Teacher Efficacy Scale .....	72
10.	Pearson Correlations Across Scales of the Two Efficacy Measures .....	74
11.	Means and Standard Deviations for the Classroom Observation Form .....	77
12.	Time and Rater ANCOVA for the Classroom Observation Form .....	78
13.	Regression Analysis of Teacher Efficacy Scales .....	80
14.	Statistically Significant Regression Correlations Between Efficacy Scales and Sub-Categories of the Classroom Observation Form Post-Observations .....	82
15.	Pearson Correlations Between the Three Student Learning Pre-Test Measures .....	87
16.	Regression Analysis of Student Learning Measures with Efficacy Scales .....	88
17.	Means and Standard Deviations of Student Learning Measures by Group .....	92
18.	Repeated Measures Analysis of Variance for Experimental and Comparison Groups .....	93

19.	Means and Standard Deviations for Low, Middle, High and Gifted Groups .....	95
20.	Multivariate Analysis of Variance for Gifted, High, Middle and Low Groups .....	96
21.	Means and Standard Deviations for the Experimental and Comparison Groups .....	100
22.	Repeated Measures Analysis of Variance for Teacher Efficacy Scales .....	101

## LIST OF FIGURES

1. Low, middle, high, and gifted group results on criterion-referenced tests ..... 97
2. Achievement group performance post-hoc test results with open-box connecting scores that are not statistically different from each other ..... 98

## ABSTRACT

The purposes of this study were to explore the relationships between teacher efficacy and changes in teacher behavior and student learning espoused by the standards-based reform movement and to examine the impact of a curriculum innovation on student learning and teacher efficacy. The study was designed to target sophisticated pedagogical behaviors associated with setting high learner expectations.

The context for the research was a federally funded project to develop and implement model lessons in elementary social studies in an urban setting. The sample was comprised of 25 experimental and 17 comparison teachers. Instrumentation included two measures of teacher efficacy, two measures of teacher behavior, and three measures of student learning.

Findings in regard to teacher efficacy were very limited. A factor analysis of the Social Studies Teacher Efficacy Scale uncovered a third factor dealing with lack of impact with difficult students that appeared to be distinct from perceptions of general efficacy, but this adapted instrument only accounted for 41% of the variance. No correlations between measures of teacher efficacy and total teacher behaviors on the observation instruments were detected, although inconsistent correlations occurred with some of the sub-categories. Weak to mild negative correlations were found between two of the sub-scales of the subject-specific efficacy measure and two of the measures of student learning. Pre- and post-test scores on efficacy did not change.

Findings in regard to the curriculum innovation were more promising. Both teachers and external observers reported a significant increase ( $p < .01$ ) in total behaviors and on four sub-categories related to educational reform expectations. Teachers self-reported higher levels of behaviors than observers. Significant gains ( $p < .01$ ) on all three measures of student learning accrued, but no differences emerged between groups. An examination of the performance of

gifted, high, middle, and low achievement students from the experimental sample only showed differences by group and measure.

The study confirmed that the measurement of teacher efficacy is complex and current instrumentation weak. There was evidence that certain dimensions of the construct may be related to specific categories of teacher behavior dealing with reform expectations, but no clear pattern emerged. Although there was tentative evidence that teacher behaviors were positively impacted by the introduction of the new curriculum, these changes appeared too shallow to affect student learning. In spite of incorporating key features from the change literature into the project design, many teachers had difficulty applying these lessons in the classroom and the overall implementation during the pilot phase was limited.

LINDA D. AVERY

PROGRAM IN EDUCATIONAL POLICY, PLANNING, AND LEADERSHIP

THE COLLEGE OF WILLIAM AND MARY IN VIRGINIA



**TEACHER EFFICACY AND BEHAVIOR:  
THEIR RELATIONSHIP AND IMPACT ON STUDENT LEARNING**

## CHAPTER I

### Statement of the Problem

#### Introduction

The standards-based educational reform movement in this country has been instrumental in underscoring the need for curriculum development work (Cohen & Spillane, 1993; O'Day & Smith, 1993). This movement advocates the articulation of challenging content standards which become the basis for the alignment of policies, practices, and resources directed toward student proficiency in meeting or exceeding the standards (McLaughlin & Shepard, 1995). At the federal and state governance levels, these standards provide the foundation for establishing policies and distributing resources across districts and schools. At the district level, the standards facilitate the design of curriculum frameworks to guide curriculum and staff development initiatives. At the classroom level, the content standards are translated into curriculum resources and instructional units that form the substance of the student's formal educational experience.

In order to improve the quality of education, the quality of the curriculum must be upgraded, and teachers must acquire the requisite skills to deliver this enriched curriculum (Shields & Knapp, 1997; USDOE/OERI, 1993). Such pedagogical skills need to be cultivated through continuous staff development cycles (Fullen, 1991), with ongoing support and technical assistance provided beyond the initial training event. Staff development that embeds pedagogical skill enhancement in the core subject matter areas is preferable to staff development that isolates the pedagogical skills from content (Shulman, 1987). This puts curriculum implementation at the center of meaningful staff development work. While improved student learning is the desired outcome of the systemic reform platform, the ability of the educational system to enhance teacher behaviors in implementing powerful curriculum experiences is the primary vehicle to

secure that outcome (Guskey, 1994; Guskey & Sparks, 1996). Thus, in order for student outcomes to change, teacher behavior must change. Teachers must continually learn and apply new knowledge and skills as they model the way for students to prepare for the demands of a future which will require adaptation to constant technological advancements in a globally interdependent economy.

This process of moving teachers from exposure to application to ownership must take into account what we know about how individuals and organizations change (Ahme, 1994; Senge, 1990). The metaphor of schools as “learning organizations” aptly conveys the idea that systemic change is a continuous process; the answer to one question leads to more questions (Sergiovanni & Starratt, 1993). Curriculum development and concomitant staff development are ongoing functions, but the work must be recursive, not reiterative. In fact, one of the failures of our educational enterprise has been the recycling of old ideas in new bottles without due regard for the philosophical and conceptual orientations that need to steer our course (Alexander, Murphy, & Woods, 1996; Fullen, 1993).

One of our understandings about change is that at the individual level it involves risk (Fullen, 1991). Individuals exist in a given comfort zone, where they have developed adaptations to the demands of their environments. Whether these adaptations are positive or negative is not the issue; the individual has the security of knowing the terrain. When the potential for change is introduced, individuals must confront their fears of the unknown.

Psychological factors are powerful determinants of one’s ability to make successful changes in one’s behavior (Maddux, 1995). Self-efficacy, a key construct in social cognitive theory, has been used to explain and predict the individual’s capacity to make changes (Bandura, 1977, 1997). Teacher efficacy, the application of this construct to the field of education

(Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998) may help us to understand the extent to which teachers adapt and respond to the demands of the reform agenda.

### Conceptual Framework

The psychological lens through which teacher behavior is being considered is based on social cognitive theory (Maddux, 1995). Social cognitive theory suggests that it is the human capacity for self-regulation and self-reflection that enables us to make choices about how we will respond to our environment. These “physiological and experiential forces interact to determine behavior and provide it with tremendous plasticity” (Maddux, 1995, p. 5).

The construct of self-efficacy, perhaps most advanced by the work of Bandura (1997), relates to the human capacity to exercise control. Through the mechanism of personal agency, people make causal contributions to their own psychosocial functioning, and among the mechanisms, “none is more central or pervasive than beliefs of personal efficacy” (Bandura, 1997, p. 2). Perceived self-efficacy is defined by Bandura (1997) as a belief “in one’s capabilities to organize and execute the courses of action required to produce given attainments” (p. 3); perhaps the best synonym is belief in one’s competence within a situational context.

Furthermore, self efficacy theory proposes distinctions among *agents*, *means*, and *ends*. Perceived self-efficacy, which refers to the *agent-means* relationship, i.e., perceptions of competence, is often separated from outcome expectancy, which deals with the *means-end* relationship, or perceptions of contingency. This distinction is clarified further in Chapter II.

One of the constructs that has shaped our investigations of the dynamics of educational reform is that of teacher efficacy. Teacher efficacy refers to “a teacher’s belief or conviction that he or she can influence how well students learn” (Guskey, 1987, p.41). Again, teachers’ perceptions of efficacy are distinct from their actual abilities to attain specific outcomes, but

there is clearly a reciprocal relationship between perceptions of competence and attendant performance outcomes. In fact, research has demonstrated that teacher efficacy is correlated with student learning (Ashton, 1985) and that teachers with high teacher efficacy evidence dispositions and behaviors that distinguish them from teachers with lower teacher efficacy (Gibson & Dembo, 1984).

### **Purpose and Research Questions**

The purpose of this study is explore the relationship between teacher efficacy and changes in teacher behavior espoused by the standards-based reform movement. These changes focus on setting more challenging academic standards, providing depth over breadth, and teaching for reasoning, problem-solving, and metacognition. Since the benchmark for successful staff development initiatives that promote teacher change is an improvement in student learning as a result of that change, the relationship between teacher efficacy and student learning will also be explored. The third component will investigate the impact of a curriculum innovation on student learning. The final component of the study will examine the extent to which teacher efficacy is impacted by a curriculum innovation.

Specifically, this research will address four questions:

1. What is the relationship between teacher efficacy and changes in teacher behavior?
2. What is the relationship between teacher efficacy and changes in student learning?
3. What is the impact of a curriculum innovation in social studies on student learning?
4. What is the impact of a curriculum innovation in social studies on teacher efficacy?

### Rationale for the Study

There are a number of areas in which relationships among teacher efficacy, teacher behavior, and student learning have not been sufficiently mapped. One of these areas is the relationship between teacher efficacy and *changes* in teacher behavior. Furthermore, what is the relationship between teacher self-report data and data gathered by external observers in the classroom? According to a summary analysis by Ross (1995), only one study documented evidence of teacher behavior change through the use of teacher observations, but that study was limited to 14 teachers (Stein & Wang, 1988).

A second area to which this study attends is the type of teacher behavior being investigated. Most studies have focused on teacher behaviors related to general classroom management practices or narrowly defined instructional modifications (Tschannen-Moran et al., 1998). However, curriculum reform requires significant pedagogical shifts in order to promote depth in learning through the application of constructivist techniques. The popular phrase for capturing such a shift is moving from the “sage on the stage to the guide on the side.” The learning theory that underlies this shift recognizes that strategies that allow meaning to be constructed by the students themselves provide more powerful and longer-lasting learning than strategies that rely solely on lecture, drill, and recitation. Do teachers’ acquisition of strategies to develop critical and creative thinking skills within content domains impact their sense of efficacy? Does the successful application of metacognitive strategies enhance efficacy? We know that using diagnostic/prescriptive techniques to help students increase their mathematical competence does appear to impact teacher efficacy (Guskey, 1984), but can more integrated approaches combining advanced content, higher order thinking, and concept learning show commensurate impacts on this variable? In other words, as a curriculum innovation requires

more sophisticated levels of teacher performance, do the same relationships between teacher efficacy and teacher behavior exist?

A third area in which additional research was warranted relates to the level of specificity at which the construct of teacher efficacy should be assessed. Most studies have examined the construct globally (Tschannen-Moran et al. 1998), but several forays have been made into specific curriculum areas, including chemistry, mathematics, language arts, and elementary science (Riggs & Enochs, 1990; Rubeck & Enochs, 1991; Curda, 1997). No previous research was uncovered for the discipline of social studies.

Furthermore, what is the relationship between the measurement of teacher efficacy globally and its measurement in social studies- specific curriculum areas? What dimensions of the construct are impacted by curriculum interventions in a specific discipline area? These inquiries became the basis for designing this study.

### Definition of Terms

The terms defined below are used throughout this study. Instruments that were used to define operationally several of these terms as they relate to the study questions are described in Chapter 3.

Culturally diverse students. This term refers to children who are “culturally different from the mainstream culture” and who are at-risk for not having their capabilities recognized (VanTassel-Baska et al., 1991, p. 3).

Curriculum. Curriculum is defined as “a set of planned experiences for a target population” (VanTassel-Baska et al., 1988). The curriculum evolves from a scope and sequence delineating the instructional outcomes scaffolded across the K-12 system and includes such key elements as specific learner outcomes, teaching/learning activities, resources, instructional

strategies, management/grouping techniques, and assessment protocols (VanTassel-Baska, 1992a).

**Curriculum innovation.** According to Senge (1990), an idea becomes an innovation when it is able to be “replicated reliably on a meaningful scale at practical costs” (p. 6) The innovation in this study involved the early stages of curriculum development and implementation in order to create a product that met the aforementioned specifications. This product consisted of 10 -12 pilot lessons incorporating instructional models aligned with the Integrated Curriculum Model (VanTassel-Baska, 1995). VanTassel-Baska’s work is targeted to the high ability learner but has been shown to have application in the regular classroom setting (VanTassel-Baska, Bass, Ries, Poland, & Avery, 1998). These pilot lessons were in social studies for grade levels two, four, and seven. Although the complete units were designed to provide 25 lessons distributed over 40-50 hours of instruction, teachers implementing the selected pilot lessons were only required to commit to 25 hours of instructional time. The pilot lessons included the specification of materials, strategies, and assessment protocols. In order to familiarize teachers with the pedagogical models, embedded in the lessons, a two day staff development workshop was held, and follow up assistance provided to teachers upon request. As part of this follow up, teachers were given opportunities to have lessons demonstrated in their classroom.

**Curriculum reform.** Curriculum reform is a platform of the standards-based educational reform movement (McLaughlin & Shepard, 1995) which calls for the setting of high standards of learning, the provision of challenging curriculum and instructional strategies to meet the standards, and assessment strategies to measure the accomplishment of the standards (Cohen & Spillane, 1993). At the classroom level, key elements of curriculum reform include having a curriculum which is meaning-based, incorporates higher order thinking skills, emphasizes inter-



and intra-disciplinary connections, provides opportunities for metacognition, develops habits of mind, promote active learning, fosters the use of technology, and uses authentic assessment strategies (O'Day & Smith, 1993; VanTassel-Baska, 1992a).

**Economically disadvantaged students.** This term refers to “children reared in homes and environments characterized by limited financial resources and educational tradition” (Frasier, 1993, p. 685). These children are often children of color, particularly African-American and Hispanic, whose families are disproportionately represented among the poorer socio-economic class (Ford, 1996).

**Gifted student.** This terms is traditionally defined as including children and youth with outstanding talent who “perform or show the potential for performing at remarkably high levels of accomplishment when compared with others of their age, experience, or environment” (USDOE/OERI, 1993, p. 26). The group of students identified as high ability learners tends to be a broader classification, referring to the top 20-25% of students in a typical general education classroom.

**Staff development.** Staff development is defined as “those processes that improve the job-related knowledge, skills, or attitudes of school employees” (cited in Moye, 1997, p. 6).

**Student learning.** Learning is defined as a “change in the cognitive structure, or in the way of perceiving events and giving meaning to them” (cited in Taba, 1962, p. 81). For purposes of this research, student learning will be examined in regard to three dimensions: (a) concept learning, (b) critical thinking, and (c) social studies content.

**Teacher behavior.** Teacher behavior is defined as classroom-based instructional behavior largely focusing on the deployment of strategies that promote higher-order thinking, problem-solving, and metacognitive skills in students. These skills are representative of what teachers

should be doing in response to standards-based educational reform (VanTassel-Baska, 1995).

**Teacher efficacy.** Teacher efficacy is defined as a “teacher’s belief or conviction that he or she can influence how well students learn (Guskey, 1987, p. 41). Personal teacher efficacy (PTE) relates to perceptions of perceived competence; general teacher efficacy relates to outcome expectations (Tschannen-Moran et al., 1998).

### **Synopsis of Methodology**

The first two research questions relied on correlational design procedures, and the next two questions used a quasi-experimental design, since random assignment was not feasible. The unit of analysis for the research questions was the group, in this case representing classrooms.

Data sources included teachers, students, and outside observers. Teacher efficacy was measured using two instruments, one a measure of global teacher efficacy, the other a subject matter-specific instrument. Teacher behavior was measured using a classroom observation form; one version of the form was completed as a self-report form by the teachers themselves (Teacher Self-Report Inventory), and the other version (Classroom Observation Form) was completed by trained external observers. Student learning was measured using a subject-specific, locally generated, criterion-referenced measure, as well as two performance measures, one related to concept learning and the other to critical thinking.

The Classroom Observation Form developed through the study was piloted to ensure adequate levels of reliability. Validity was determined by a rating by content experts. Other instruments, already in use, have been determined to have adequate indices of reliability and validity.

Data analysis techniques involved multiple regression for the first two research questions; the third and fourth question were analyzed using a repeated measures analysis of variance in

order to treat time as an independent variable.

### Significance of the Research

This research was designed to extend our understanding of the variable of teacher efficacy in some interesting directions. First, it explored the relationship between *changes* in teacher behavior and teacher efficacy. This issue of change was very important. Most studies of teacher efficacy have correlated it with specific teaching behaviors. Only a few prior studies examined changes in behavior brought about by an intervention, and of these only one was found which used observational data to validate teacher change.

Second, this study examined some important dimensions of teacher behavior, extending the research into pedagogical practices aligned with standards-based educational reform. Of particular interest were teacher behaviors related to challenging learners, promoting critical thinking and problem-solving abilities, and enhancing metacognitive skills. These teacher behaviors were deemed critical in advancing the reform agenda in American schools.

Third, this study examined whether a curriculum innovation that was aligned with key elements of curriculum reform influenced teacher efficacy. If teacher efficacy is associated with student performance, it is important to continue investigations of the pathways that enhance or impede it.

Fourth, this study used multiple measures of teacher efficacy to investigate the relationship between global and specific dimensions of the construct. This attempted to strengthen our understanding of the contextual dimensions of the construct.

These contributions were intended to advance our understanding of the relationships among teacher efficacy, curriculum implementation, and student learning. This information was important in guiding various aspects of the development of the innovation as it progressed from

the pilot to the field-test stage.

### Context for the Research

This research capitalized on the award of a federal grant to develop and field-test model curriculum units in the social studies. The collaborators on the grant were the Center for Gifted Education at the College of William and Mary and Norfolk Public Schools. Specifically, seven elementary and middle schools in the Norfolk district participated in the curriculum implementation initiative. As a result, the student target population for the curriculum was primarily urban, with six of the seven schools having a majority of students on free- or reduced-lunch status. The curriculum was, therefore, designed to respond to the cognitive and affective needs of the disadvantaged, culturally diverse high ability learner.

### Limitations and Delimitations

The major limitations of this study were sample size and lack of random assignment of teachers to treatment or comparison classes. Although the study initially attracted 28 experimental teachers, these teachers were distributed over three grade levels, with each grade getting different pilot lessons, united primarily by the instructional models embedded in them. This restricted grade level analysis of the data.

Another concern related to sample size was the limited variability in responses from the teacher sample on the measures of teacher efficacy. This was particularly problematic for the correlation with teacher behavior, because no comparison teachers were observed. With such a small sample size, variability was somewhat constrained and without variability, correlations were less likely to emerge.

The issue of lack of random assignment also restricted the potential for attribution of causality if the study had found treatment effects from the impact of the curriculum innovation

on student learning. Since the study was largely exploratory at this stage, all findings were treated with caution.

The study was delimited by the selection of a specific urban site with a high incidence of economically disadvantaged and minority students. The timeframe for the study reduced the opportunity to reveal impacts that might have emerged after more cycles of implementation occurred. A third delimitation was in the selection of instruments to measure student achievement. The use of a locally designed criterion-referenced instrument did not specifically address the objectives of the new units and may have masked important learning that accrued to students as a result of the intervention. Additional measures of student learning, developed by an external evaluator, were intended to compensate for this delimitation.

## CHAPTER II

### Review of the Literature

#### Introduction

The introduction of a new curriculum into the educational environment, particularly one that is grounded in pedagogical practices advocated by educational reform, must take into account what is known about change theory and practice. Our understanding of change has evolved from describing it as a process that is straightforward, linear, and incremental (Havelock, 1969) to one that is complex, recursive, and ongoing (Senge, 1990). Within this new paradigm, schools are characterized as “learning organizations,” committed to a central vision of quality education that is executed through a process of continual improvement at the individual (personal mastery) and group (team learning) levels (Senge, 1990). In order to increase a sense of teacher ownership, many schools have decentralized decision making, thus making “every person a change agent” (Fullen, 1993, p. 39). Such responsibility is an integral element of a systems perspective, in which the whole is composed of interlocking, interdependent parts that must function harmoniously to solve current and future problems.

#### Bridging Federal and Local Levels of the System

The curriculum innovation that is the foundation of this research bridged both federal and local levels of this system. The federal level provided funding for the innovation in keeping with its role of stimulating change through demonstration grants (Wirt & Kirst, 1997), a role that has been the subject of national debate in the educational research community. Despite large-scale investments and high expectations, researchers have lamented that few innovations have made a dent in solving national educational problems (Havelock, 1977; Vinovskis, 1993). Klein (1993) attributed the lack of impact beyond the target site to a variety of factors, including failure to (a)

identify the unique features of a context, (b) advertise promising solutions to relevant consumers, and (c) collect important feedback for revision. His reasons highlighted the heightened regard for the variable of “context” in understanding how to advance the educational reform agenda.

At the local level, it is the educators themselves who impact the success or failure of the innovation. Alexander et al. (1996) cited two potential hypotheses to explain why so many promising innovations fade into oblivion. The first, captioned “Doing what we know,” dealt with the human tendency to tackle or respond to what is easy rather than hard; the second, “Knowing about what we do,” focused on the failure of educators to have a rich understanding of the intended innovation or sufficient knowledge of the literature or research bases underlying such innovations. In investigating this second hypothesis, the researchers uncovered evidence that today’s educators have a paucity of “knowledge of the people, movements, and writings that underlie many recurring educational innovations” (p. 31), which makes them slaves to the interpretations of others and vulnerable to the latest bandwagon. This research gave further credence to Fullen’s assertion (1991) that one of the greatest problems in education is not resistance to innovation, but the “fragmentation, overload, and incoherence resulting from the uncritical acceptance of too many different innovations” (p. 197).

### **Making Innovation Work**

The process of moving from initiation to implementation to institutionalization requires a balance between individual needs and concerns and organizational goals and demands (Ahrne, 1994; Fullen, 1991). At the organizational level, Guskey (1990) identified five guidelines to aid school leaders in their efforts to integrate innovations: (a) all innovative strategies should share common goals and premises; (b) no single innovative strategy can do everything; (c) innovative strategies should complement each other; (d) innovative strategies need to be adapted to

individual classroom and building conditions; and (e) results from a combination of well-conceived innovative strategies are likely to have greater impact than from any single strategy.

While Guskey's suggestions are mostly aimed at administrators, other educators have made suggestions that impact classroom practice. Alexander et al. (1996) offered four key directives: (a) seek principled understanding, (b) teach more about less, (c) aim for rooted relevance, and (d) reward reflection over revolution. Wong (1997) articulated three foci for promoting teacher adoption of research-based instruction: (a) address teachers' overarching conceptions of what teaching is about, (b) enhance teachers' subject area knowledge, and (c) attend to the research on the impact of student learning on teacher efficacy.

Borko, Mayfield, Marion, Flexer, and Cumbo (1997) also identified factors that facilitated the change process in promoting the use of performance assessments in mathematics with a group of third grade teachers in three schools. Although the scope of their innovation was narrower than the curriculum innovation in this study, it, too, was tied to classroom level reform, and as a result, their findings, presented as themes, were pertinent to this study. Their five themes identified the importance of (a) situating the change process in the contexts in which the ideas will be implemented; (b) using group discussions as a tool for the social construction of new ideas and practices; (c) using staff development personnel or other "experts" to introduce new ideas based on teachers' levels of interest, understanding, and skill; (d) recognizing that incompatible beliefs held by teachers must be challenged before new practices will be adopted; and (e) recognizing that "time is a major obstacle to changing classroom practice" (Borko et al., 1997, p. 272).

#### **Features of the Innovation Related to Change**

It was within this context of understanding how to effectuate change that the curriculum



innovation was positioned. Features of the introduction of the innovation into the environment based on the change literature included the following: (a) use of teacher volunteers (Fullen, 1991), (b) use of teams to support collaboration (Senge, 1990), (c) selection of model lessons (less is more) to illustrate new pedagogical models (Alexander et al., 1996), (d) allowance for classroom and building adaptation (Guskey, 1990), (e) provision of support through technical assistance in the classroom context (Borko et al., 1997), (f) securement of administrative commitment at the building and central office levels (Fullen, 1993), (g) alignment of the curriculum innovation with newly developed curriculum frameworks and standards of learning (Guskey, 1990; McLaughlin & Shepard, 1995).

It was through this lens of attending to the change process that the curriculum innovation was understood as a microcosm of the larger educational reform agenda. Against this backdrop of educational reform, additional aspects of the literature were reviewed that underscored the particular research questions of interest.

### **Organization of Review of the Literature**

This review of the literature focused on current theory and research across four major strands. The four strands were (a) teacher efficacy and its relationship to teacher behavior and student learning; (b) teacher change through staff development and transfer of learning; (c) curriculum improvements for high ability learners; and (d) the special needs of disadvantaged and culturally diverse gifted learners. A summary of key points in the literature around each of these themes follows.

#### **Strand I: Teacher Efficacy and Its Relationship to Teacher Behavior and Student Learning**

Teacher efficacy as a construct for understanding teaching effectiveness has been in the literature for several decades (Tschannen-Moran et al., 1998). This strand of the literature review

focuses on the theoretical framework for the construct and its measurement history, research findings related to its relationship to teacher behavior, and its relationship to student learning. This section ends with a discussion of the implications of the strand for the present study.

### Conceptual Orientation, Research Tradition, and Measurement Implications

The first component of this strand of the literature on teacher efficacy provides an overview of social cognitive theory, the construct of self-efficacy, a definition of teacher efficacy, a summary of the research approaches measuring teacher efficacy, and the implications of this measurement history.

Conceptual orientation and definition. Teacher efficacy is a construct derived from social cognitive theory. Social cognitive theory proposes that personal factors in the form of cognition, affect, and biological events interact with the environment to adapt human behavior in a tripartite, reciprocal relationship. People's mental capacities for self-reflection and self-regulation enable them to be "active shapers of their environments rather than simply passive reactors to them" (Maddux, 1995, p. 4). Social cognitive theory encompasses a large set of factors that operate as regulators and motivators, such as self-efficacy, cognitive processes, affective processes, and motivational processes (Bandura, 1997). Self-efficacy, however, "occupies a pivotal role in social cognitive theory because it acts upon the other classes of determinants" of behavior (Bandura, 1997, p. 35). Self-efficacy influences behavior in four ways (Bandura, 1997; Ross, 1995). Through *cognitive processes*, high self-efficacy contributes to the adoption of loftier goals, more goal commitment, and higher outcome expectancies. Through *motivational processes*, high self-efficacy results in self-attribution of success or failure. Through *affective processes*, high self-efficacy promotes coping strategies such as positive self-talk, and through *selection processes*, self-efficacy influences the choice of activities or environments an

individual makes. As a result of these influences, self-efficacy beliefs are strong determinants “of the level of accomplishment that individuals finally attain” (Pajares, 1996, p. 545).

Self-efficacy is a regulatory mechanism that relates to people’s beliefs about their capabilities to “organize and execute the courses of action required to produce given attainments” (Bandura, 1997, p. 3). According to Bandura (1997), perceived self-efficacy “is not a set of skills one has but a belief about what one can do under different sets of conditions with whatever skills one possesses” (p. 37). Self-efficacy theory distinguishes between competence, or agent-means relationships (I can execute the actions), and contingency, or means-ends relationships (the actions will attain certain outcomes) (Tschannen-Moran et al., 1998). Furthermore, self-efficacy is distinct from other conceptions of self, such as self-concept and self-esteem, in that it is specific to a particular task or set of tasks (Bandura, 1997).

Bandura (1997) has acknowledged that an individual’s efficacy belief system is “not an omnibus trait” but is “a differentiated set of self-beliefs linked to distinct realms of functioning” (p. 36). Teacher efficacy, therefore, is the application of self-efficacy theory to professional educators in the context of the educational system. The definition of the term itself has undergone some evolution, but generally refers to “the extent to which teachers believe their efforts will have a positive effect on student achievement” (Ross, 1995, p. 228). Guskey (1987), for instance, defines it as a “teacher’s belief or conviction that he or she can influence how well students learn, even those who may be difficult or unmotivated” (p. 41). The construct has been investigated with practicing teachers, both new and experienced, and preservice teachers (Ross, 1995; Tschannen-Moran et al., 1998).

Research traditions and measurement of the construct. The research on teacher efficacy has grown out of two major traditions in the literature and has been well summarized by

Tschannen-Moran et al. (1998). One strand emerged from the theoretical base of Rotter's work on locus of control, i.e., whether "control of reinforcement lay within (teachers) themselves or in the environment" (cited in Tschannen-Moran et al., 1998, p. 202). Teachers with a high level of internal locus of control believed they could strongly impact student performance and motivation. The second and prevailing conceptual strand emerged from the work of Bandura (1977, 1997) described previously. Efficacy beliefs influence how much effort people put forth, how long they will persist in the face of obstacles, how resilient they are in dealing with failures, and how much stress they experience. Bandura (1977) distinguished between efficacy beliefs, which deal with perceptions of future performance, and outcome expectancies, which deal with the perceived consequences of performance.

Both strands of research have resulted in an understanding of teacher efficacy as a multi-dimensional construct with the most common two factors identified as personal teaching efficacy (PTE) and general teaching efficacy (GTE). There is a reasonable consensus across the different research strands that PTE has to do with "one's own feelings of competence as a teacher," but the meaning of the second factor, GTE, is still in question (Tschannen-Moran et al., 1998, p. 223). According to Tschannen-Moran et al. (1998) some researchers have described the second factor as external influences, similar to external locus-of-control, while others align it with Bandura's notion of outcome expectancy, having reasoned "that what teachers in general could accomplish was the outcome an individual teacher could expect from his or her own teaching" (p. 223).

Research treating GTE as an outcome expectancy has relied heavily on an instrument developed by Gibson and Dembo (1984), the Teacher Efficacy Scale (TES). However, a recent re-examination of the factors on this scale concluded that GTE, as measured on this instrument, more accurately reflects the impact of elements that lie beyond the direct control of teachers, i.e.,

external influences such as peers or home (Guskey & Passaro, 1994). In other words, when the items on TES were adjusted to control for confounding variables, the alignment of factors validated Rotter's rather than Bandura's conceptual framework. Rather than reflecting outcome expectancies, the factors grouped around external rather than internal influences.

Furthermore, interpretation of the two factors has always been confounded by additional context variables. For instance, in early studies, Rose and Medway (1981) and Guskey (1981) reported an interaction between internal and external locus of control and positive and negative outcomes for students. These studies suggested that teachers assumed more responsibility for positive outcomes than for negative ones. Other studies examined the impact of additional context variables, such as ability level of students and scope of outcome (individual versus group), on teacher efficacy responses. Cooper, Burger, and Seymour (1979) reported that teachers felt they had less control over low ability students and were less able to influence how these students learned. Guskey (1987) found that when the performance outcome was negative, teachers distinguished between individuals and groups. Teachers expressed less responsibility for failure with a single student than for a group or entire class.

Another issue that affects measurement of teacher efficacy is the level at which the construct is assessed. Most scales that measure teacher efficacy treat the construct broadly (Gibson & Dembo, 1984; Guskey, 1981, 1987; Rose & Medway, 1981). For example, Bandura's own approach to the development of a teacher efficacy instrument identified seven subscales based on tasks teachers are required to perform, such as instructing students, disciplining students, involving parents, and influencing resource decisions (Bandura, 1997). His measure attempted to provide a multifaceted picture of teachers' efficacy beliefs in relation to the variety of roles they assume. However, even within the instructional tasks of teachers, "teachers' sense

of instructional efficacy is not necessarily uniform across different subjects” (Bandura, 1997, p. 243).

Other researchers have modified the Gibson and Dembo instrument to explore teachers’ sense of efficacy within particular curriculum areas. Riggs and Enochs (1990) explored the construct in relation to the teaching of science. Rubeck and Enochs (1991) distinguished chemistry teaching efficacy from science teaching efficacy. Curda (1997) investigated an instrument designed to focus on assessing efficacy for reading and mathematics instruction. No evidence was found in the literature examining social studies teaching efficacy per se.

Although the issue of generality versus specificity is still unresolved, researchers agreed that context (including content areas, student characteristics, etc.) is important in understanding and measuring this construct (Bandura, 1997; Tschannen-Moran et al., 1998). Ross (1995) suggested that there was evidence for teacher efficacy to be treated as a relatively stable core surrounded by dimensions that fluctuate in response to the characteristics of specific teaching assignments. This argues for approaches that tap both a global sense of teacher efficacy and the more specific context in which it is being investigated.

Implications for measurement of teacher efficacy. Teacher efficacy is a complex construct that has been explored with evolving instrumentation. There is uncertainty about the level of specificity at which the construct should be measured (Bandura, 1997; Tschannen-Moran et al., 1998), confusion regarding the alignment of instrumentation with theoretical orientations (Guskey and Passaro, 1994), and widespread inconsistency in the selection of instruments (Ross, 1995; Tschannen-Moran et al., 1998). Nevertheless, there is much agreement that teacher efficacy is a powerful construct that predicts both classroom behavior and student outcomes (Ross, 1995; Tschannen-Moran et al., 1998). According to Ross (1995), research on teacher efficacy has entered its adolescence: “enormous strides have been made in the last five years in

our understanding of the construct and its influence” in education (p. 244).

### **Teacher Efficacy and Teacher Behavior**

This component of Strand I examines the relationship between teacher efficacy and teacher behavior. Studies which have focused on teacher behavior in regard to instructional and classroom management practices, rather than teacher beliefs, are the primary focus.

**Research on the relationship between teacher efficacy and teacher behavior.** While the research base on the relationship between teacher efficacy and teacher behavior is not extensive, the findings across studies have, for the most part, found similar trends. Rose and Medway (1981) investigated the relationship between teacher efficacy and classroom behavior based on a Teacher Locus of Control Scale (TLC) that they developed and on observations by outside observers during mathematics instruction. They found that high efficacy teachers in low socio-economic status (SES) schools gave fewer disciplinary commands to students, and high efficacy teachers in high SES schools called on nonvolunteers more frequently. A second component of the study, conducted with a different sample of teachers participating in inservice training, found that high efficacy scores correlated positively with use of a variety of classroom materials, use of learning centers, and provision of group projects.

Gibson and Dembo (1984) used their own scale to investigate relationships between teacher efficacy and classroom behavior through observation. Teachers with high efficacy devoted more time to academic activities, offered more guidance to non-achievers, were less likely to criticize students following incorrect responses, and were more likely to divide the class for small group instruction. Schunk (1995), in summarizing findings from a study by Ashton and Webb, echoed similar ideas: teachers with high efficacy were more likely to have a positive classroom environment, support students’ ideas, and check on student progress in learning.

Payne's study (1994) of teachers, identified as significant or not by student ratings, found countervailing patterns. She selected a subsample of eight teachers and observed their classrooms. Her own ratings were triangulated with ratings by external auditors of the audio-tapes of these classrooms. As a result of this process, she identified behaviors of teachers who were deemed more significant by their students. These behaviors included utilization of lesson plans, ability to give feedback and illustrate lessons, effective use of reinforcement, use of higher-order questioning strategies, and use of lessons personalized to students' abilities and interests. However, using the Teacher Efficacy Scale (Gibson & Dembo, 1984), she found an inverse relationship between teacher significance and personal teacher efficacy (PTE) based on the 35 teachers who volunteered for the larger study. Clearly, the small number of teachers in the subsample and the qualitative approach to data analysis limited generalizability of her results.

In examining classroom practices of science teachers based on self-report, Riggs and Enochs (1990) noted that teachers with high personal science teaching efficacy (PSTE) spent more time teaching science. Additional research by Riggs (1995) reported on teachers involved in a year-long training program in science education. Those with low PSTE spent less time teaching science, used a text-based approach, were rated weak by site observers, and made fewer positive changes in their beliefs about how children learn science.

Research on the impact of staff development on teacher efficacy. In addition to correlational studies that have examined relationships between teacher efficacy and teacher behavior in the classroom, six quasi-experimental studies investigated whether skill development through inservice education impacts teacher efficacy (Ross, 1995). According to a summary analysis prepared by Ross (1995), three of these studies suffered from design flaws that made the findings questionable. A fourth study, conducted by Cousins, Maynes, and Ross himself (cited in



Ross, 1995), had inconsistent findings across groups on new versus experienced teachers. The final two studies, discussed below, provided more credible data.

Guskey (1984) used three measures, his Responsibility for Student Achievement scale (RSA), a teaching self-concept scale, and an attitudes toward teaching scale, on a pre- and post-test basis on two groups of intermediate and high school teachers (N=117). One group (n=52) participated in inservice training on mastery learning (diagnosis, feedback, and correction) with the other serving as the control. After training, the experimental teachers, having agreed to teach two sections of the same course, were asked to randomly select one of their classes in which to use the new techniques but to withhold the techniques in the other. Student outcomes were measured at the end of the course, using multiple indices to classify students as having positive, negative, or no change. "Teachers who had experienced a positive change in the learning outcomes of their students (n=34) felt more responsible for both positive and negative student outcomes and expressed more positive attitudes toward teaching than did the other groups of teachers" (Guskey, 1984, p. 252). Guskey suggested that, in the absence of positive change in instructional effectiveness (i.e., student outcome improvements), inservice training and practice implementation had little effect on teachers' affective characteristics.

Surprisingly, Guskey (1984) also found that evidence of teachers' own teaching effectiveness resulted in a reduction in teacher confidence. He explained this by suggesting that when teachers realize that there is a better way to teach, their confidence is temporarily shaken.

According to Ross (1995), the "most persuasive evidence for a causal connection among in-service training, teacher efficacy, and student achievement" was presented in a study by Stein and Wang. Stein and Wang (1988) observed changes in the teacher efficacy of 14 teachers over three school terms during the implementation of a specific innovation. Changes in teacher

practice occurred between terms one and two, preceding changes in teacher efficacy that emerged between terms two and three. Teachers who changed the most in terms of their use of the new practices showed the greatest growth in teacher efficacy.

Although the data are extremely limited, they tend to show that teacher efficacy is fairly stable unless the implementation of new practice results in improvement in student learning. These studies assessed teacher efficacy as a global construct and focused on an innovation that was narrowly defined. However, both studies spoke to the delayed effect on teacher efficacy of student outcomes.

### Teacher Efficacy and Student Learning

There is a substantive research base tying teacher efficacy to student learning. In fact, the early RAND studies which investigated teacher efficacy through the use of two items (linked to personal and general teaching efficacy) found strong relationships between teacher efficacy and reading and mathematics achievement (Armor et al., 1976; Berman, McLaughlin, Bass, Pauly, & Zellman, 1977). In a study of basic skills teachers at four secondary schools, Ashton and Webb (1986) reported that general teaching efficacy (the factor related to the means/end relationship or outcome expectancy) increased the explained variance in math achievement scores (measured on the Metropolitan Achievement Test) by 24%. Personal teaching efficacy (the factor related to perceptions of the agent/means relationship) explained 46% of language achievement.

Other studies, using the Teacher Efficacy Scale developed by Gibson and Dembo (1984), found similar relationships. Moore and Esselman (1992) found correlations with math skill for second and fifth graders on the Iowa Test of Basic Skills. Anderson, Greene, and Loewen (1988) validated a positive relationship between PTE and third graders' achievement on the Canadian Achievement Tests, reporting also that teacher efficacy was a stronger predictor with younger

than with older students. Ross (1992) showed that both PTE and GTE related to higher levels of achievement on the Ontario Assessment Pool. Watson's (1991) study of third graders also found significant relations between teacher efficacy orientation and student reading and mathematics achievement in both majority Black and majority White schools.

Ross (1995) pointed out that the association of teacher efficacy with achievement is "based exclusively on correlational data. Even though the findings are consistent, the empirical relationship could be the product of an unexamined third variable" (p. 230). He further summarized the two arguments undergirding the claim that teacher efficacy contributes to student outcomes. The first argument is that teachers with high teacher efficacy set high standards for themselves; the second is that they set high standards for students. His analysis of the research suggested that both alternatives are plausible. His third argument considered the possibility of a reverse correlation, i.e., that student achievement increases teachers' assessment of their competence. Ross concluded by stating that "the relationship between teacher efficacy and student outcomes is to some unknown degree reciprocal" (p. 230).

In addition to the research on individual teacher efficacy, there have been some attempts to examine "collective teacher efficacy." Bandura (1997) aggregated teachers' beliefs about their school's capacity to foster academic attainment. He discovered that teachers' sense of collective efficacy varied across grade level and subjects, but he found that teachers' beliefs in the school's efficacy as a whole was just as predictive of school performance as teachers' beliefs in their own efficacy. This research has opened new avenues for the investigation of teacher efficacy in relation to systemic reform initiatives (Bandura, 1997).

### **Implications of the Research on Teacher Efficacy, Teacher Behavior, and Student Learning**

**Research on student achievement has focused heavily on linking standardized**

achievement gains in mathematics and reading to global measures of teacher efficacy. Although teacher efficacy has been shown to correlate with student achievement, there is variability in the factors (i.e., PTE or GTE) that account for this alignment. Furthermore, incipient evidence suggests that the pathway for influencing this relationship involves first changing teacher behavior to a level that impacts student performance; then evidence of student learning gains appears to impact level of teacher efficacy (Gersten & Guskey, 1985; Guskey, 1982, 1984, 1986; Guskey & Sparks, 1996; Ross, 1998). Teacher training in the absence of follow up implementation that supports improved student outcomes appears to have little impact on teacher self-efficacy (Guskey, 1984). Additional studies need to be undertaken to validate this relationship and extend its application across disciplines (Curda, 1997; Riggs & Enochs, 1990).

A second implication from this strand of the research review is that teacher behaviors and changes in such behaviors in relation to teacher efficacy have been defined somewhat narrowly. Mastery learning approaches, use of teacher feedback, time spent on task, and approaches to student discipline are examples of classroom behavior that have been investigated either through correlational or quasi-experimental research (Guskey, 1984; Payne, 1994; Rose & Medway, 1981). The relationship between teacher efficacy and teacher behaviors or changes in behavior associated with deeper curricular and pedagogical initiatives has not been substantively addressed. Many opportunities abound for defining additional elements of classroom practice designed to support higher standards of learning and for investigating the impact of such behavioral changes on the construct of teacher efficacy. Such elements as use of critical and creative thinking strategies, metacognition, delivery of advanced content, and attention to concepts are specific examples of teacher instructional behavior yet to be correlated with teacher efficacy.

## **Strand II: Teacher Change through Staff Development and Transfer of Learning**

The review of this strand of the literature was grounded in the cognitive psychological perspective on learning. There are three major assumptions that are central to this perspective that have application to studies of learning to teach (Borko & Putman, 1996). They are (a) the central role of knowledge in thinking, acting, and learning; (b) learning as an active, constructive process; and (c) knowledge and learning as situated in contexts and cultures. According to Borko and Putman (1996), strategies for staff development programs that take advantage of what we know about the process of learning to teach are as follows:

1. Make existing knowledge and beliefs the object of reflection and scrutiny.
2. Ground learning opportunities in the teaching of specific subject matter domains.
3. Treat teachers as learners and create experiential learning environments.
4. Ground learning in classroom practice.
5. Provide sustained time and support for reflection, collaboration, and continued learning.

The following sections of this strand of the literature review describe the centrality of student learning as a focus for successful staff development efforts, provide additional guidelines for making staff development effective, and summarize information on the transfer of learning from the staff development experience to the classroom. The section ends with a discussion of the implications of this strand for the research being conducted.

### **The Centrality of Student Learning in Staff Development Efforts**

Increasing student learning is the goal of the standards-based reform agenda (McLaughlin & Shepard, 1995) and the benchmark against which educational innovations must be evaluated (Pogrow, 1998a). Furthermore, research has shown that top-down, structural, and/or governance

reforms do not necessarily translate into classroom level changes (Guskey & Peterson, 1996; Shields & Knapp, 1997; VanTassel-Baska et al., 1996a). Without impact at the classroom level, the fundamental mission of change efforts, to improve student learning, remains unaddressed.

Furthermore, what is seen as the linchpin in getting traction at the classroom level is staff development linked to student learning improvement (Moye, 1997; Shulman, 1987; Sparks, 1994). Guskey and Sparks (1996) proposed a model for illustrating the relationship between staff development and student learning. In their model, the quality of staff development, affected by content, process, and context characteristics, drives teacher knowledge and practice, this, in turn, impacts student learning. While there are also other influences that directly and indirectly impact student learning (such as school policies and parent practices), staff development is the primary vehicle for changing teacher behavior. Guskey and Peterson (1996) advised policy makers to “invest in high-quality professional development and make significant changes in the way these activities are planned, organized, and carried out” (p. 13). This recommendation recognizes that staff development linked to student learning is an important determinant of educational reform and that traditional approaches to staff development have been too narrowly construed (Sparks, 1995).

### **Making Staff Development Effective**

Guskey (1994) offered guidelines for making professional development effective. Examples of his suggestions included the following: (a) recognize that change is both an individual and organizational process, (b) think big but start small, (c) work in teams to maintain support, (d) include procedures for feedback on results, and (e) provide continued follow-up, support, and pressure. While these ideas were essentially applications of the change literature to the function of staff development, they reenforced the relevance of this knowledge base to the

teaching and learning enterprise and the centrality of staff development as the vehicle for impacting student growth. Other research on staff development spoke more specifically to the content and structure of effective approaches.

Effective staff development work must be predicated on grounding pedagogical knowledge in the content-domain, rather than encouraging the separation of process and content (Gardner, 1991; Shulman, 1987). This understanding of staff development links the teacher and the curriculum, instead of treating improvements in instruction separately from improvements in content, materials, and assessment. In writing about staff development, Sparks (1995) alleged that educators have failed to use “systems thinking”; attempts are often made to enhance instruction without seeing it as an element of a larger curriculum framework that includes curriculum, instruction, and assessment. He recommended content-specific staff development that addresses deeper forms of content knowledge keyed to instructional strategies most effective in that discipline.

Sparks (1995) also noted that staff development should provide a model of the desired practices, “because example is such a powerful teacher” (p. 166). For instance, if teachers are taught that students need to construct meaning for themselves in order to attain deeper understanding of knowledge, then teachers need to experience constructivist techniques in their inservice program. Simon and Schifter (1993) reported successful outcomes using this maxim. A constructivist-oriented inservice program provided to mathematics teachers resulted in more classroom emphasis on conceptual understanding and less on computational skills. Student outcomes indicated that attitudes toward math improved, even as their standardized test scores were maintained. While this example does not meet Pogrow’s (1998a) criterion for an exemplary program, it does show the power of modeling the skills that are addressed in staff development.

In other research, Showers, Joyce, and Bennett (1987) identified the structural components for effective staff development. The components they cited include demonstration of the strategy, practice, feedback, and follow-up in the classroom. Cognitive coaching and follow up were recommended to overcome discouragement or apathy during the transition process. According to Joyce and Showers (1995), coached teachers were more likely to practice new skills, develop greater expertise in skill areas, and deploy the skills more effectively than non-coached peers. Wong (1997) also suggested that the presence of colleagues who can serve as models or mentors encourages reluctant teachers to initiate the use of new strategies/approaches.

#### Teacher Transfer of Learning into Practice

The last component of the Showers et al. model introduced the concept of transfer. It is not enough for teachers to learn new ideas and behaviors through staff development experiences if they are not translated into practice in the classroom. Moye (1997) reviewed the literature on transfer of learning into classroom practice and identified four conditions that provide support for transfer: (a) having a strong, positive school culture, (b) incorporating the elements of effective training, (c) providing opportunities for coaching or follow up, and (d) recognizing the link between increases in student learning and teacher efficacy. Guskey (1986) proposed a model for the process of teacher change which posited that teachers' attitudes and beliefs (including teacher efficacy beliefs) change only after they have evidence that new practices increase student learning. The sequence of elements in the model suggests that changes in teachers' beliefs support inferences that teachers made successful attempts to translate new learning into actual practice and that the student outcomes were favorable.

#### Implications of this Strand of the Research

Cognitive psychology offers a powerful framework for understanding and structuring the



process that helps teachers learn to teach. Because improvements in student learning are at the heart of educational reform, staff development must be clearly linked to improving student learning (Guskey, 1994). In order to link teacher development to student learning, attention must be paid to the content and processes of the training. If we want students to have better conceptual knowledge, teachers need to be trained on concepts as well as the pedagogical processes in teaching to and monitoring the learning of the concepts. Inservice that is content-focused and embeds pedagogy within the relevant discipline is more effective than training that teaches skills independent of subject matter. The structure of teacher training is also important and extends beyond the initial training event through the provision of follow up and support during the implementation period. The enacted curriculum (Ball & Cohen, 1996), which brings together the teacher's beliefs, knowledge, and expertise in fashioning or executing an educative experience for a group of learners, is the point of ultimate accountability.

### **Strand III: Curriculum Improvements for High Ability Learners**

The previous section of this literature review focused on the pivotal role of staff development in changing teacher behaviors to promote student learning. This section speaks to the importance of curriculum as the centerpiece of the staff development process. Teachers become better teachers by teaching the curriculum better or by strengthening the curriculum that is taught. In both cases, improved teaching practice goes hand-in-glove with curriculum reform.

### **Curriculum Reform in the Context of Educational Reform Initiatives**

According to Cohen and Spillane (1993), "instructional guidance in the United States has been inconsistent and diffuse" (p. 59). While state and federal governments have made many efforts to improve instruction, their policies "rarely make broad or close contact with instruction" (Cohen and Spillane, 1993, p. 46). Consequently, the standards-based reform initiatives have

stepped into this breach by calling for the setting of high content standards, with accountability tied to the demonstration of student proficiency with these standards (McLaughlin & Shepard, 1995). The articulation of national content standards in mathematics, science, language arts, and the social sciences has provided guidance to state and local educators as well as sparking public discourse on curriculum standards (McLaughlin & Shepard, 1995). It has also led to “using curriculum as a basis for talent development for all learners” (VanTassel-Baska, 1995, p. 98).

Studies investigating the elements of successful educational reform efforts have acknowledged the criticality of reform at the classroom level. In a study of exemplary middle schools, VanTassel-Baska, Hammett-Hall, & Bailey (1996) found insufficient evidence that structural change resulted in classroom level changes. Similarly, Shields and Knapp (1997) found that school-based reform does not guarantee improved student outcomes. Consequently, they urged schools to focus explicitly on particular aspects of curriculum, while targeting related professional development concerns. Cawelti’s (1995) study of high school restructuring identified the setting of performance standards, the use of interdisciplinary curriculum, and the integration of technology into the classroom as three of the seven critical elements associated with improved learning environments. A study by Phillips (1997) documented that middle schools with an academic philosophy, rather than a communitarian philosophy, were more likely to show student learning gains in mathematics. Clearly, higher student expectations and challenging curricula are at the heart of effective classroom reform in this country.

#### **The National Perspective on Curriculum for Gifted and Talented Students**

While the national debate about the need for general educational reform was launched with the publication of A Nation At Risk (United States Department of Education/ National Commission on Excellence in Education, 1983), it was not until the publication of National

Excellence: A Case for Developing America's Talent (USDOE/OERI, 1993) that a national perspective on the education of gifted and talented children within the context of this larger reform agenda was articulated. This report echoed the clarion call for setting “challenging curriculum standards” for all students, including the high ability learner, and encouraged the provision “of more and better opportunities for top students to learn advanced material and move at their own pace” (USDOE/OERI, 1993, p. 2). Citing concerns about the performance of our best students both domestically on the National Assessment of Educational Progress (USDOE/OERI, 1990) and internationally on math and science comparisons across industrialized countries, the report urged policy makers and practitioners to attend to this “quiet crisis” in American education. The report (USDOE/OERI, 1993) indicated that gifted students in this country “are offered a less rigorous curriculum, read fewer demanding books, do less homework, and enter the work force or postsecondary education less well prepared than top students in other countries” (p. 5). Furthermore, “most of them continue to spend time in school working well below their capabilities” (p. 5).

Of the five recommendations offered in the report to address the problem, two had direct implications for curriculum and instruction at the classroom level. The first of these called for the establishment of challenging curriculum standards, suggesting that performance standards in the core subject areas should be sufficiently high to challenge talented students. While this recommendation was targeted to state and local political entities, it has clear ramifications for the specific standards of learning adopted by school districts and addressed by classroom teachers. These standards are often the basis of curriculum frameworks which guide the selection of materials, instructional strategies, and assessment protocols (Cohen & Spillane, 1993). If these frameworks do not sufficiently address depth of learning, mastery of higher order thinking skills,

and conceptual knowledge needed to frame interdisciplinary connections, then it is unlikely that the curriculum will bridge these gaps.

The second recommendation in the National Excellence report (USDOE/OERI, 1993) called for the establishment of high-level learning opportunities for high ability students. This recommendation emphasized the importance of accelerating the rate of learning of gifted students and the provision of in-depth work in the core curriculum. This focus on the substance of the gifted child's education, rather than on the organizational arrangement used to deliver the "gifted program," reinforces the pivotal role of curriculum in driving educational excellence. Although the report was cognizant of the alarming lack of services for gifted children in the regular classroom (Westberg, Archambault, Dobyns, & Salvan, 1993), improvements in the nature of curriculum and instruction were clearly posited as the primary vehicles for addressing this problem.

This tack, while politically prudent, was also supported by the findings of a study of grouping practices. In a meta-analysis on the impact of grouping practices on student performance, Kulik and Kulik (1992) reported that ability grouping of high-end learners without curricular modification was ineffective. This does not mean that homogeneous grouping of gifted learners is undesirable. In fact, it is often recommended by proponents of gifted education; but without addressing the quality of the curriculum and instruction that is offered, grouping is not a panacea for gifted education service delivery (Benbow, 1998; VanTassel-Baska, 1992b).

Most students of high ability spend most of their school day in the regular classroom, particularly at the elementary level (VanTassel-Baska, Avery, & Hall, 1997; Westberg et al., 1993). Therefore, curriculum reform that elevates the quality of curriculum and instruction in the regular classroom clearly aligns with the interests of the field of gifted education. The question

then becomes, to where does one turn to find promising curriculum models with exemplary pedagogical practices?

Because gifted education has been credited with developing services around higher expectations and challenging curricula (Callahan, 1996; Tomlinson & Callahan, 1992), the field has much to contribute to an understanding of general educational curriculum reform. Such curricular enhancements as focusing on higher order thinking skills, incorporating creative problem-solving, and addressing metacognition were touted in the field of gifted education long before entering the mainstream (Tomlinson & Callahan, 1992). To this end, knowledge of effective curriculum for gifted learners may provide guidance for how to strengthen general education classroom practice.

#### Curriculum Design for High Ability Learners

The nature of what constitutes the appropriate curriculum for gifted and talented students has been part of the literature base in the field for most of the last century. Leta Stetter Hollingworth, one of the pioneers in the gifted education field and widely regarded for her work in curriculum design and development, emphasized the importance of enriched and accelerated educational experiences for gifted youngsters as early as the 1920s (Passow, 1990). Key elements of her curriculum development work included attention to interdisciplinary connections, use of biography to study great contributors to society, and an emphasis on inquiry and discussion as pedagogical staples (VanTassel-Baska & Brown, in press).

For the past two decades, gifted programs in this country have been influenced by two competing program orientations which have shaped much of the discourse on the talent development process (VanTassel-Baska & Brown, in press). In one camp are the proponents of radical acceleration who have targeted the highly gifted (top 1-3% on standardized tests) and

have shown that individual variability in this group of students is profoundly diverse and that our threshold for understanding the academic capacities of these youngsters has been set far too low (Stanley, 1976). This body of research has validated the need for higher standards of learning and more attention to individualization.

The second philosophical orientation, also widely endorsed, has targeted a larger segment of the population and focused on classroom compacting and enrichment through small group and individual project work (Renzulli, 1975). This model has offered a vision of gifted education that is more inclusive in identification and tailored to implementation in the general education classroom. Although the original designer of this model has suggested that it provides a detailed blueprint for school improvement (Renzulli, 1996), there is no research showing that the curriculum materials per se are effective.

While these program orientations have had ramifications for curriculum development efforts, they are somewhat flawed in serving as the template for curriculum design work. The radical accelerants focused on a narrow band of gifted youngsters whose interests were served by shortening their exposure to the K-12 program; the enrichment specialists emphasized individual project work and process skills. As a result, other curriculum models warrant attention.

### Review of Curriculum Models in Gifted Education

VanTassel-Baska and Brown (in press) offered a definition of the term curriculum model based on the identification of five components: (a) a framework for curriculum design and development, (b) transferability and use in all content areas, (c) K-12 applicability, (d) applicability across schools and grouping settings, and (e) the incorporation of differentiated features for gifted and talented learners. They further identified 15 criteria for assessing curriculum development models in gifted education and applied the criteria to 12 distinct models.

The result of the culling process was that only six models appeared to have research data documenting their effectiveness with gifted learners. One of these models, the Integrated Curriculum Model (ICM) (VanTassel-Baska, 1995) met 12 of the 15 criteria, earmarking it as a valuable curriculum model for use in curriculum development initiatives. Features which distinguished it from most of the other models were its (a) alignment with national curriculum standards, (b) utilization by almost one hundred school districts in the country, and (c) emergent database assessing longitudinal student impacts. Curriculum units developed by the College of William and Mary utilizing this model were awarded the National Association for Gifted Children's exemplary curriculum award in 1997, 1998, and 1999.

#### Research on the Integrated Curriculum Model

The Integrated Curriculum Model was specifically developed for use with high ability learners (VanTassel-Baska & Brown, in press), and research has demonstrated the effectiveness for this population of curriculum units in language arts and science designed under the aegis of this model. Research conducted in the area of language arts showed significant growth in literary analysis and interpretation, persuasive writing, and linguistic competency for students in grades four through six exposed to the unit entitled Autobiographies: Personal Odysseys of Change, using a quasi-experimental design with statistical controls to equate for the differences between the experimental and comparison classes (VanTassel-Baska, Johnson, Hughes, & Boyce, 1996). Student growth in science process reasoning skills, the result of 20-25 hours of instruction in the unit entitled Acid, Acid Everywhere, also proved statistically significant with an effect size of 1.30 based on data gathered from 45 experimental and 17 comparison classes (VanTassel-Baska et al., 1998). Additional research is being collected to validate the effectiveness of other units based on this model. However, research on the model is limited to units designed for elementary

and middle school students.

### **Overview of the Integrated Curriculum Model**

The Integrated Curriculum Model (VanTassel-Baska, 1995) was developed in response to the perceived lack of “a comprehensive and cohesive curriculum framework that uses good curriculum design, that considers the features of the disciplines under study, and that sufficiently differentiates for talented students” (p. 99). Research has shown that there is as much variability in intelligence within the population identified as gifted (Gagne, 1993) as there is within the population falling between one standard deviation above and below the mean. Therefore, the notion of differentiation is critical to curriculum design. In order for sufficient differentiation to occur, the curriculum must provide different levels of challenge for the students exposed to it.

VanTassel-Baska (1995) explained that the ICM was grounded in the research on learning. The embedding of higher order thinking skills in subject matter enhances transfer of learning (Perkins & Saloman, 1989). Teaching the concepts of a discipline promotes longer-term retention (Marzano, 1992). The model is also consistent with research by Pogrow (1998b) on learning gains achieved with Elementary and Secondary Education Act, Title I students. His study found that teaching for understanding required sophisticated interventions applied in a systematic, sustained, and intensive way.

The ICM has three interrelated curriculum components (VanTassel-Baska, 1995):

1.     **Advanced content dimension focused on new learning and validated through the use of diagnostic/prescriptive approaches;**
2.     **Process/product dimension focused on higher order thinking skills and the utilization of advanced reasoning in a generative way;**
3.     **Issues/themes dimension focused on real world applications and theoretical**



modeling within and across areas of study.

These elements complement such characteristics as precocity, intensity, and complexity of gifted learners (VanTassel-Baska, 1995), but when above average children in the general education classroom have been exposed to the curriculum, they have also responded favorably to the level of challenge presented (Myrtle, 1997).

#### Alignment with Curriculum Reform Design Elements

The curriculum template afforded by the ICM also complements key design aspects of curriculum reform articulated by O'Day and Smith (1993). For instance, the curriculum is meaning-based, incorporates higher order thinking skills, emphasizes inter- and intra-disciplinary connections, provides opportunities for metacognition, develops habits of mind, promotes active learning, capitalizes on the availability of technology, and uses authentic assessment (VanTassel-Baska, 1995). These attributes speak to the strong conceptual grounding that scaffolded the development of the model.

#### Implications of this Strand of the Research

Curriculum reform has emerged as the catalyst to improve learning outcomes. Since curriculum in its broadest sense serves as the nexus between the teacher and the student, it also must be the focus of staff development efforts (Shields & Knapp, 1997). Key aspects of standards-based curriculum reform relate to the promulgation of challenging content standards, the design and delivery of curriculum that address these standards, and the provision of accountability for student learning in accordance with the standards (McLaughlin & Shepard, 1995). In order to ensure that students have maximum opportunities for learning, the curriculum must emphasize advanced content, higher order thinking/reasoning skills, and important concepts. These dimensions must be integrated to ensure that students are able to construct

meaning and to translate understanding into generative products (VanTassel-Baska, 1995).

The Integrated Curriculum Model developed by VanTassel-Baska (1995) for gifted learners provides a viable template for actualizing curriculum development work. Although the model has not yet been applied to the discipline of social studies, the model's adaptation across both the humanities and the sciences bodes well for the extension of the model into a new curriculum development initiative.

#### Strand IV: Economically Disadvantaged and Culturally Diverse Gifted Students

This section of the literature review examines the research on economically disadvantaged and culturally diverse gifted learners. Gifted students who are culturally different from the mainstream and/or economically disadvantaged have been defined as at-risk populations (VanTassel-Baska, Patton, & Prillaman, 1991). In particular, Hispanic-Americans and African-Americans have been under-represented in gifted programs in this country (Baldwin, 1985). As the demographics of this country's student population shift from predominantly White to larger percentages of children of color, our understanding of the how to identify and serve gifted children within the larger cultural milieu must also change (Maker, 1996).

The National Excellence report (USDOE/OERI, 1993) highlighted the national interest in increasing "opportunities for disadvantaged and minority students with exceptional talent to participate in advanced learning experiences" (p. 28). The report admonished schools to eliminate barriers that preclude the identification and involvement of promising gifted and talented learners from poor neighborhoods and minority populations. Since many children in urban school settings fall into the demographic categories cited by this report, it is important to review the literature on educational issues relevant to their learning needs.

### Under-Representation in Gifted Programs

A considerable literature base exists documenting the under-representation of economically disadvantaged children and selected minority populations in gifted education programs in this country, constituting a “large reservoir of untapped and under-developed talent” (Passow, 1991, p. 1). Data from the National Educational Longitudinal Study showed that only 9% of students in gifted and talented education were in the bottom quartile in family income, while 47% were from the top financial quartile (cited in USDOE/OERI, 1993). An earlier study by VanTassel-Baska et al. (1991) cited similar findings. For instance, although students from low-income families comprised 20% of the student population, they made up only 4% of students who performed at the highest levels on standardized tests (95<sup>th</sup> percentile or above). Also, high school seniors from disadvantaged families were less than half as likely as more advantaged seniors to participate in gifted programs.

Because of the disproportionately high incidence of racial and ethnic groups in the lower socio-economic strata, African-Americans are particularly negatively impacted. Recent research indicated that 48% of Black children under age 13 lived in homes with incomes below the poverty level (Ford, 1996). Although African-Americans represented 16% of the public school population, they constituted 27% of all students classified as having trainable mental retardation or serious emotional disturbance, 30% of all students expelled, and 31% of those who have received corporal punishment (Ford, 1996). Margolin (1994) reported that only 8% of identified gifted students were Black, a figure consistent with findings cited in National Excellence (USDOE/ OERI, 1993). Furthermore, Black females outnumbered Black males in gifted programs by a ratio of two to one, and Black males were the group that consistently scored the lowest on standardized tests of achievement (Ford, 1996).

Efforts to reverse this situation by modifying identification protocols have met with only modest success. In fact, in Project Mandala, researchers attempting to test an identification model using a profile approach, found that non-traditional measures were no more effective than traditional measures in locating the population of interest (Ward et al., 1992). This speaks to the complexity of using identification modifications as the solution for reaching under-represented gifted populations.

The under-representation of these students in traditional gifted education programs adds impetus to the importance of strengthening curriculum and instruction in the general education classroom. In fact, the history of exclusion of these children from the ranks of the identified gifted has created a faction in the field of gifted which denounces homogeneous grouping of gifted students in both self-contained and pull-out programs (Margolin, 1994; Sapon-Shevin, 1995). This perspective is consonant with the current preference for inclusion of special populations, including the gifted (Culross, 1997; Delisle, 1994), in the regular classroom. Curriculum reform that differentiates for advanced learners in the general education classroom is likely to reach more high ability students from low-income families and more culturally diverse gifted populations.

#### **Learning Needs of Economically Disadvantaged and Culturally Diverse Gifted Students**

While the research base on the under-representation of low-income and culturally diverse gifted learners was fairly substantial, there were fewer studies that reported on educational practices addressing their needs. VanTassel-Baska et al. (1991) synthesized the generic features of interventions that work well with disadvantaged and culturally different gifted students: (a) early and systematic attention to their needs, (b) parental and family involvement in their educational program, (c) use of effective schools' strategies, (d) experiential and "hands-on"

learning approaches, (e) activities that allow for self-expression, (f) mentors and role models, (g) counseling that addresses cultural values and facilitates talent development, and (h) building on the strengths and differential learning styles of at-risk learners. Maker and Schiever (1989) presaged many of these same elements in their earlier synthesis of the literature on programs and curricula for cultural and ethnic minorities. Their summary included the following: (a) plan the curriculum based on students' strengths, (b) provide for the development of basic skills and other abilities that the student may lack, (c) consider differences as positives, (d) arrange for mentors, and (e) create and maintain a classroom with a multicultural emphasis.

Frasier (1993) cited several examples of U.S. programs that had success in working with disadvantaged and minority gifted populations, including the Skills Reenforcement Project (SRP) and the Program of Assessment, Diagnosis, and Instruction (PADI). Tomlinson, Callahan, and Lelli (1997) found that multicultural emphases, language immersion, use of manipulatives, participation in mentorships, and family outreach were essential for the success of low-income and/or minority primary-aged gifted students in Project START (Support to Affirm Rising Talent).

Other researchers have focused on the larger cultural context for understanding the needs of specific minority groups. Clasen (1992) and Ogbu (1994) have both commented on the alienation experienced by involuntary minorities struggling to juggle the expectations of conflicting worlds. For instance, African-American gifted students who demonstrate academic aptitude can be rejected by their Black peers for "acting White." As a result, these students must develop coping mechanisms in order to straddle the value systems of competing allegiances (Patton & Baytops, 1995). Such demands create additional psychological stressors in minority gifted populations that need psycho-social support (Ford-Harris, Schuerger, & Harris, 1991).

In studying the needs of at-risk gifted adolescents, Olszewski-Kubilius, Grant, and Seibert (1994) noted that programs that address only one aspect of the child's life, such as the provision of an appropriate academic program, have "little hope of having long-term impact" (p. 23). This underscores the need for parent involvement in the child's education. Parent involvement has been found to have a positive effect for students when it provides appropriate modeling, reinforcement, and even direct instruction within a holistic context of family, school, and community (Epstein, 1995; Hoover-Dempsey & Sandler, 1995). Epstein (1995) found that developing school, family, and community partnerships improved school programs and school climate and provided a basis for connecting families to services in the community. Such refrains are also echoed in the literature on resilience (Werner & Smith, 1992), which advocated for multifaceted, systemic interventions in the home, school, and community.

#### **Implications of This Strand of the Research**

The under-representation of economically disadvantaged and culturally diverse students in gifted programs and services provides fodder for the argument to elevate the quality of curriculum and instruction in the general education classroom (Sapon-Shevin, 1995). Research on the learning needs of many of these students, particularly from African-American and Hispanic populations, offered some guidance about program and curriculum interventions that have worked (Maker & Schiever, 1989; VanTassel-Baska et al., 1991). Key features of successful programs included early identification and sustained intervention, experiential and hands-on learning, and use of curricula with multicultural emphases that allow for student choice (Maker & Schiever, 1989; Tomlinson et al., 1997; VanTassel-Baska et al., 1991). In addition, an awareness of the importance of parental involvement, or support by a mentor who encourages academic achievement, permeated this segment of the literature (Olszewski-Kubilius et al., 1994;

VanTassel-Baska et al., 1991).

In urban settings with large numbers of minority and economically disadvantaged students, curriculum interventions must be particularly sensitive to issues of multi-culturalism and learning styles. By incorporating knowledge from this strand of the literature into both the staff and curriculum development components of the project, there was greater likelihood for the innovation to be successful in gaining teacher and student acceptance in order to accomplish its aims.

### Implications from the Overall Review of the Literature

Because the context for this study was a federally funded project awarded to the Center for Gifted Education at the College of William and Mary that involved collaboration with a local school district, it was important to attend to the literature on change theory as a backdrop to the study. This literature indicated that the federal role in funding educational innovations has been somewhat controversial and that it has been difficult for programs funded with federal dollars to solve national problems (Vinovskis, 1993). Nevertheless, there is an expectation that demonstration grant programs should be designed to facilitate replication so that innovations can be disseminated in a cost effective manner to other similar settings.

Juxtaposed against these federal expectations is the desire at the local level to have an innovation that is tailored to the particular needs of the context. Local educators, both administrators and teachers, are most interested in addressing local problems with solutions that fit their particular environment. While both the federal and local educators may share common interests, there is some level of creative tension in fashioning responses that meet both sets of expectations or demands. In order to increase the chances of success in modulating these demands, it is important to be cognizant of the literature on effecting change at the organizational

as well as the personal levels (Fullen, 1993). Therefore, many strategies taken from this literature were employed in the design and implementation of the curriculum innovation that was a catalyst for this study.

In addition, the review of the literature examined four strands which converged on the research questions being proposed. The first strand investigated the construct of teacher efficacy and the research linking the construct to teacher behavior and student learning, primarily within the theoretical orientation of social cognitive theory (Bandura, 1997). While this strand of the review demonstrated that there appears to be a significant correlation between high teacher efficacy and certain positive teacher behaviors associated with classroom instruction (Ross, 1995; Tschannen-Moran et al., 1998), there are additional aspects of teachers' instructional competence that have not been studied. Several of these aspects relate to the impact of the standards-based reform movement in calling for more challenging curriculum and the incumbent use of more sophisticated pedagogy (Cohen & Spillane, 1993; McLaughlin & Shepard, 1995). Therefore, the relationship between teacher efficacy and changes in teachers' competence in delivering new curricula through more complex instructional strategies was a target for exploration in this study.

Furthermore, since improved student outcomes appear to be one catalyst for impacting the construct of teacher efficacy, the correlation between student outcomes and teacher efficacy might serve as a proxy for the validation of teacher change (Guskey, 1984). Guskey (1984) and Stein and Wang (1988) have shown that teacher efficacy can change when there is evidence that changes in teacher behavior produce learning gains in students. Therefore, the measurement of student learning gains incorporated into the research design was intended to inform our understanding of the dynamics of the primary relationship.



Since virtually no research has examined teacher efficacy within the context of social studies curriculum and instruction (Curda, 1997; Riggs & Enochs, 1990), this study has begun to fill that void. While there is some confusion about what exactly the instruments assessing self-efficacy actually measure (Guskey & Passaro, 1994), there is agreement that contexts appear to have some impact on this construct (Bandura, 1997; Tschannen-Moran et al., 1998). Therefore, the nature of the teacher behaviors being assessed and the student outcome measures impacted by the behaviors need to be domain specific to make the inferences as tight as possible.

The second strand of the literature review examined the research on teacher change, staff development, and transfer of learning. Since staff development is the vehicle for initiating and supporting behavioral changes in classroom teachers (Guskey, 1986; Sparks, 1994), an understanding of the knowledge base in this area was necessary. The literature clearly established that staff development is a recursive process that requires continuous, ongoing support for teachers beyond the initial training experience (Fullen, 1991; Senge, 1990). The foremost purpose of staff development is to elevate teaching practice to impact student learning (Guskey, 1994; Sparks, 1995). Therefore, staff development is viewed as a critical component of school improvement initiatives (McLaughlin & Shepard, 1995).

Student outcomes should serve as the criterion for determining the effectiveness of staff development efforts (Guskey & Sparks, 1996). In order to impact student learning, staff development must be tied to curriculum reform (Shields & Knapp, 1998). Content knowledge must be integrated with procedural knowledge within specific disciplines so that teachers are trained on the relevant pedagogy to support the delivery of the new curriculum (Shulman, 1987). In addition, staff development should unfold with attention to the structure of adult learning that involves demonstration, practice, feedback, and follow-up (Showers et al., 1987).

The third strand in the review of the literature focused on curriculum improvements for high ability learners as the centerpiece of staff development efforts. The national vision articulated by educators of the gifted is consonant with content-based standards reform (USDOE/OERI, 1993). Both call for the setting of high standards and the provision of challenging learning experiences to help students excel (McLaughlin & Shepard, 1995; VanTassel-Baska, 1995). Because the field of gifted education has had a strong focus on curriculum almost since its inception and has pioneered many of the popular strategies for promoting higher-level thinking skills, the field offers rich resources to support and inform educational reform at the classroom level (Callahan, 1996; Tomlinson & Callahan, 1992). Of particular relevance to this research was a curriculum framework proposed by VanTassel-Baska (1995) which integrates content, process/product, and concept dimensions within discipline areas and links them to appropriate pedagogical and assessment strategies. This framework guided the curriculum development effort in social studies that shaped the design and delivery of curricular and instructional innovations.

The fourth strand in the literature review identified research related to the needs of the economically disadvantaged and/or culturally diverse high ability learner. Since the demographics of U. S. schools are rapidly changing and since our urban districts are particularly reflective of children of color from poor neighborhoods, issues and concerns regarding the needs of this population should be considered (Maker, 1996; USDOE/OERI, 1993). This literature base supported the general education classroom as the locus for reaching this group of learners (Delisle, 1994; Westberg et al., 1993). Their absence from traditional programs and services for identified gifted students, coupled with the fact that even most gifted students spend the preponderance of their school day in the regular classroom (at least at the elementary level)

(Passow, 1991; VanTassel-Baska et al., 1997), speaks to the regular classroom setting as the point of greatest leverage (Sapon-Shevin, 1995). In addition, the literature in this area addressed the need to respond to the whole child through the acknowledgment of multicultural perspectives, the use of experiential learning strategies, and outreach to the home and family (Maker & Schiever, 1989; VanTassel-Baska et al., 1991). These considerations were incorporated into the curriculum innovation in order to optimize the match between the social studies program changes and the learners.

## CHAPTER III

### Methodology

This study was designed to contribute to the literature on teacher efficacy in relationship to teacher behavior and student learning and to examine the impact of a curriculum innovation on both student learning and teacher efficacy. The research focused on the domain of social studies instruction at selected grades in both elementary and middle schools in an urban community with high concentrations of economically disadvantaged and culturally diverse students.

#### Conceptual Framework

The conceptual framework for understanding self-efficacy was based on the work of Bandura. Bandura (1977) proposed that the construct of self-efficacy is a future-oriented belief about the level of competence a person expects to display in a given situation. It deals with a perceived expectation of performance and links together the two components of *agent* and *means*. In Bandura's (1997) social cognitive theory, there is also a recognition of outcome expectancy. Outcome expectancy ties together the two elements of *means* and *ends*. In this equation, the individual (agent) perceives acting at a certain threshold of behavior (means), which impacts on the anticipation of certain results (ends). For instance, a person with low self-efficacy for driving may anticipate a clumsy performance; this performance may or may not create an expectation for an accident. Bandura (1997) asserted that outcome expectancies add little to the predictive power of efficacy measures.

When the construct of self-efficacy is applied to education, it is usually translated into measures of teacher efficacy. Most instrumentation based on Bandura's work includes items which address the two dimensions of efficacy; one dealing with personal teacher efficacy (PTE) and the other with general teacher efficacy (GTE) (Tschannen-Moran et al., 1998). These

dimensions are often interpreted as assessing the two relationships embedded in Bandura's theory: perceived efficacy and outcome expectancy.

Bandura (1997) also asserted that context is an important dimension in understanding self-efficacy. Teachers may feel efficacious working in some content areas and not others, or working with certain grade levels or student populations. Most studies of teacher efficacy have examined the construct globally, but a few studies have focused specifically on the sciences. Studies done in science have used an adaptation of the Teacher Efficacy Scale (Gibson & Dembo, 1984) for a specific content area such as elementary science or chemistry (Riggs & Enochs, 1990; Ruback & Enochs, 1991). No studies were found which examined teacher beliefs in relation to social studies teaching efficacy.

### Research Questions and Instruments

Although psychometric data are provided later in this chapter, the instruments that were used to measure each research question are described below:

1. What is the relationship between teacher efficacy and changes in teacher behavior?

For this question, teacher efficacy was measured using two instruments: a) Teacher Efficacy Scale (short-form) and b) Social Studies Teacher Efficacy Scale. The first is a measure of global teacher efficacy, and the second a subject-specific measure. Changes in teacher behavior were measured using two instruments: a) Classroom Observation Form (COF) completed by external observers, and b) Teacher Self-Report Inventory, a version of the COF that teachers themselves completed.

2. What is the relationship between teacher efficacy and changes in student learning?

For this question, teacher efficacy was measured using the same two efficacy instruments,

the Social Studies Teacher Efficacy Scale and the Teacher Efficacy Scale (Short-form). Changes in student learning were measured using the Norfolk district Criterion-Referenced Test (CRT) in Social Studies. In addition, two newly developed measures were included, a Concept Learning Test and a Critical Thinking Test, both tied to dimensions of the curriculum innovation.

3. What is the impact of a curriculum innovation in social studies on student learning?

For this question, the three measures of student learning were used: the Norfolk CRT in Social Studies, the Concept Learning measure, and the Critical Thinking Measure. In addition, data from the Teacher Demographic Questionnaire was analyzed to see if there were critical differences between the experimental and comparison teachers.

4. What is the impact of a curriculum innovation in social studies on teacher efficacy?

For this question, the two sets of efficacy scales were administered to both experimental and comparison teachers.

### Research Design

These research questions were investigated using a design drawn from the traditional quantitative research paradigm (Gall, Borg, & Gall, 1996). This research framework was selected for two reasons. First, most of the work that has been done in regards to teacher efficacy has used quantitative approaches (Tschannen-Moran et al., 1998), so the use of this methodology enabled comparisons with prior research findings. Second, the quantitative approach used group data, which masked individual results, thus promoting greater willingness on the part of teachers to participate in the study.

The first two research questions were stated in terms of relationships between key

variables, requiring a correlational research design. In order to assess correlations with *changes* in teacher behavior and student learning, the measures of these variables were given on a pre- and post-test basis. The third and fourth questions anticipated a cause and effect relationship which called for a quasi-experimental design (Gall, Borg, & Gall, 1996). This also necessitated the administration of pre- and post-measures on both the student learning instruments and the teacher efficacy scales.

### Site Selection

The study was carried out with teachers and students from seven school buildings in Norfolk Public Schools in Norfolk, Virginia. Norfolk is an urban school district with portions of its geography designated as an enterprise zone by the federal government, due to a high incidence of families living in poverty. The district is 60% African-American, with 70% of the overall student body on free or reduced lunch.

The seven schools selected for the study had significant percentages of children on free or reduced lunch, ranging from 46% to 84% . Five of the schools were elementary (K - 5); one of the schools was a middle school (6 - 8), and another was a K - 8 school. Furthermore, these seven school buildings volunteered to participate in the federally funded project awarded, under the Javits grant program, to The College of William and Mary. Criteria for the selection of the seven schools included willingness of the principal and the teachers to volunteer, absence of other major curriculum reform initiatives being undertaken concurrently, and the majority of the student body on free- or reduced lunch status. All but one of the sites met this last criterion.

In addition, specific teachers at grades two, four, and seven were encouraged to volunteer for the project based on their experience level; no first year teachers were included in the group targeted for piloting the curriculum. All other teachers at the same grade levels were asked to

serve as comparison teachers for the first wave of implementation.

The project itself involved the design, development, and field-testing of pilot curriculum lessons in social studies, with interdisciplinary connections to mathematics and technology. The lessons were drawn from units developed by the Center for Gifted Education at The College of William and Mary in collaboration with the Norfolk school district. The pilot lessons were aligned with both the state's and the Norfolk district's Standards of Learning (SOL) and were based on the Integrated Curriculum Model (VanTassel-Baska, 1995) for high ability learners.

The Deputy Superintendent for Academic Affairs and Accountability granted permission for the research in the district. Additionally, individual teachers, serving in either the experimental or comparison cohort, signed consent forms explaining the parameters of the research and their right to withdraw at any time without penalty. These forms contained language prescribed by the Human Subjects Review process at the College.

#### Description of the Intervention

The dependent variable that constituted the intervention was the introduction of 10 - 12 new curriculum lessons within the social studies content domain based on the Integrated Curriculum Model (VanTassel-Baska, 1995). Staff development to enhance teachers' abilities to deliver the new lessons by using constructivist pedagogical strategies accompanied the implementation process. Key aspects of the curriculum innovation and the staff development support are described below:

Curriculum innovation. The curriculum innovation involved the creation and implementation of 10 - 12 pilot lessons from three new instructional units in social studies, one each for grades two, four, and seven. The lessons addressed both the state and the district SOL. Teachers were required to spend at least 25 hours of instructional time delivering the lessons. If



teachers had already covered the content of the lessons during the first semester of the school year, they were permitted to adapt the pedagogical models to new content. They were also given a separate research unit tied to a problem-based learning scenario with could be taught as an addendum or in lieu of other lessons to meet the time commitments. The lessons emphasized key elements of curriculum reform articulated by O'Day and Smith (1993). Such features included being meaning-based, incorporating higher order thinking and problem-solving skills, emphasizing inter- and intra-disciplinary connections, providing opportunities for metacognition, developing habits of mind, promoting active learning, fostering the use of technology, and using authentic assessment strategies.

Staff development. The staff development model was based on a comprehensive understanding of what was needed to support teachers in implementing innovation in the classroom (Sparks, 1995). Staff development was initiated with a two day workshop for experimental teachers and focused on the new curriculum lessons and the relevant instructional strategies to enhance implementation of the lessons. In addition, staff development was supported on an ongoing basis by providing technical assistance. This assistance was available through the involvement of both the media personnel in the sites and the centralized gifted program staff. Staff from the district gifted program and graduate students from the Center for Gifted Education were on-site at least four times during the second semester to conduct observations and answer questions. In addition, teachers were told that demonstration teaching of the lessons could be done in their classroom at their request. Assistance with securing resources was also provided in several cases. Project staff and district personnel were available to answer questions raised by teachers before and during the implementation process.

Because principal support was also seen as critical to the success of classroom change (Fullen, 1991), principals were given a briefing on the project. At that time, they were asked to consider using the Classroom Observation Form as a way of recording what they saw during their formal and informal forays into the classroom. The intent of this was to help principals focus on the set of teacher behaviors that were being stressed as a way of supporting the schools' internal conversations about the implementation of the project.

### Sample

The sample for this study was originally comprised of 28 experimental and 17 comparison teachers and their respective students in the seven target schools at grades two, four, and seven. The experimental teachers volunteered for the study and the remaining teachers at each of the three grade levels agreed to serve as the comparison group. Attrition of the experimental teachers is described in Chapter IV. Experienced teachers were encouraged by building principals to volunteer for the pilot-test of the new curriculum lessons; random assignment was not feasible.

### Instrumentation

Eight instruments were used to carry out the study. Copies of all but one of the instruments is included in the Appendix. This excluded instrument is the property of the district and not available for distribution. A description of each instrument and its relevant psychometric properties are described below.

Teacher efficacy scale (short-form). This instrument provided a measure of global teacher efficacy and was adapted from the Gibson and Dembo (1984) version of the Teacher Efficacy Scale. The short-form had 10 items which were rated on a six-point scale (Woolfolk & Hoy, 1990). The personal teaching efficacy subscale had an alpha of .77 and the general teaching

efficacy subscale had an alpha of .72 (Hoy & Woolfolk, 1993). The architects of the revised scale encouraged researchers using this instrument to conduct factor analysis on their own data because the loadings had not been consistent across studies. Consequently, a factor analysis based on the study sample's responses was conducted. The results of this factor analysis are presented in Chapter IV.

Social studies teacher efficacy scale. This instrument was adapted from the Science Teaching Efficacy Beliefs Instrument (STEBI) developed by Riggs and Enochs (1990). The STEBI is composed of two scales, one measuring Outcome Expectancy and the other, Personal Science Teaching Efficacy Belief, with a total of 25 items using a Likert scale. The reliability and factor analysis of the first scale resulted in an alpha of 0.91 and item-total correlations of 0.50 and above for all items. The reliability alpha for the second factor was lower, at 0.73 with item-total correlations at 0.37 and above. Measures of construct validity were deployed, but they were not relevant to the adaptation for social studies. In order to adapt the scale for social studies, six graduate students were asked to complete the instrument. No rewording of items occurred as a result of their feedback. However, a factor analysis of the new version of the instrument was conducted based on the study sample's responses. The results of this factor analysis are presented in Chapter IV.

Classroom observation form. This instrument was developed by the Center for Gifted Education under the direction of VanTassel-Baska. Observers are first instructed to outline or script their observations. Then they are required to translate these observations into a coding format. This coding contains 40 items specifying behaviors which an observer records as observed or not observed during a 30 to 45 minute observation period. A composite score is computed by summing the items present. The instrument has nine sub-categories, and no sub-

category has less than three items. The largest sub-category has nine items. The instrument was used by graduate students, faculty, or gifted support teachers in the Norfolk district who had received training on the form. The researcher collected inter-rater reliability data from a pool of trained observers during the fall, 1998. The results of this analysis showed a median kappa of .63. Content validity data were secured by asking three experts in the field of gifted education and one instructional coordinator in general education to rate the instrument. On a scale of 1 to 3, with 3 being high, the mean rating for the instrument was 2.9 (or .96) for content validity. Furthermore, the feedback from these experts was used to make minor revisions in the form prior to actual piloting.

Teacher self-report inventory. This instrument was an adaptation of the Classroom Observation Form that was completed by the teacher at the conclusion of the observation period. It has all the same items, but teachers were not required to do a narrative recording of their instructional behavior. The instrument measures the teacher's perception of his or her execution of the behaviors on the observation form during the implementation of the lesson.

Teacher demographic questionnaire. This short instrument contained items regarding the demographic characteristics of the teachers engaged in the study. It included items such as highest level of education, number of years teaching, number of years teaching in this school, number of years teaching at this grade level, year hired into district, date of birth, and gender. The form was piloted with graduate students prior to distribution to study teachers.

Norfolk district criterion-referenced test at grades 2, 4, and 7 in social studies. These tests were developed by the district to measure student performance on the district's Standards of Learning. Psychometric data on their validity and reliability were never calculated. The pilot of the project constituted the first wide-scale implementation of this new instrument. Copies of this

instrument must be secured from the district. The scoring of the instrument was handled by the district, with student results grouped by teacher sent directly to the researcher. Although the items on the test were not well matched with the content of the curriculum innovation, the district wanted to limit the testing demands on students and teachers.

Concept learning test. The Concept Learning Test was developed by the external project evaluator and contained items which were aligned with the concept being addressed in the curriculum i.e., “systems,” and with aspects of Taba’s work (1962) on concept development. The pilot phase of the project was used to establish the psychometric properties of the instrument (Rogers, 1999). Pearson Product Moment correlation coefficients were calculated for the pre- and post-test scores on this test. Test-retest reliability was strong ( $r = >.62$ ), but “not as high as one would desire” (Rogers, 1999, p. 9). An item analysis conducted on the test showed that most items had enough ceiling for growth to be measured.

Critical thinking test. The Critical Thinking Test was developed by the external project evaluator and contained items which were aligned with the Paul model of reasoning (Paul, 1992). The pilot phase of the project was used to establish the psychometric properties of the instrument (Rogers, 1999). Pearson Product Moment correlation coefficients were calculated for the pre- and post-test scores on this test. Moderate test-retest reliability was found ( $r = .50$ ). An item analysis conducted on the test showed that most items had enough ceiling for growth to be measured.

### Data Collection

Table 1 describes the data collection framework for the instruments delineated previously.

Table 1

**Data Collection Framework**

Instrument	Experimental Teachers/Classrooms		Comparison Teachers/Classrooms	
	Pre	Post	Pre	Post
<b>Student Data</b>				
District CRT for Soc. Studies	X	X	X	X
Concept Learning Measure	X	X	X	X
Critical Thinking Measure	X	X	X	X
<b>Teacher Behavior Data</b>				
Classroom Obs. (4 ea)	X	X		
Teacher Self-Report (4 ea)	X	X		
<b>Teacher Demographic Data</b>				
Teacher Demographic Survey	X		X	
<b>Teacher Efficacy Data</b>				
Teacher Efficacy Scale (short-form) - Global measure	X	X	X	X
Social Studies Teacher Efficacy Scale- Subject-specific measure	X	X	X	X

**Data Analysis Procedures**

The primary statistical methods deployed in this research involved the use of hierarchical regression and repeated measures analysis of variance techniques (Grimm & Yarnold, 1995). The first research question was addressed by correlating data from the experimental teachers only on the two self-efficacy instruments and the two measures of classroom behavior. Each efficacy

measure had two scales so hierarchical regressions were run with each scale. In addition to running correlations with the total scores on the observation instruments, correlations were run with each of the 9 sub-categories. In order to establish correlations with *changes* in teacher behavior, each teacher's pre-observation score was entered into the regression equation first.

The second question was addressed by correlating data from the total teacher sample on the two self-efficacy instruments and the three student learning measures. Again, hierarchical regressions were run with each scale of the efficacy instruments. In order to establish correlations with *changes* in student learning, each classroom pre-test score was entered into the regression equation first. This was followed by entering the pre-test scores of the other instruments since all three measures shared some common variance.

The third question was addressed by comparing the performance of the experimental and comparison groups on the three student learning measures. A repeated measures analysis of variance was used which treated time as an independent variable. This statistical procedure was selected because it was sensitive to changes in learning over time as well as interaction effects based on group membership.

The fourth question was addressed by comparing the experimental and comparison teachers on self-efficacy changes between the pre and post-test administration of the instruments. Again, a repeated measures analysis of variance was used which treated time as an independent variable.

#### Time Frame for the Study

The study was conducted during the second semester of the 1998-99 academic year. Testing of the experimental teachers on the self-efficacy instruments occurred at the beginning of the teacher staff development event in early January. Comparison teachers were asked to

complete the instruments at the time their students were being pre-tested. Pre-implementation observations and self-reports occurred from late January through March. Post-training observations and self-reports transpired near the end of the unit implementation process which varied across sites. Student pre-tests were administered before unit implementation; student post-tests at the conclusion of the implementation process which was in late May or early June. Teacher post-testing on the efficacy instruments occurred in late May or June.

### Confidentiality and Other Ethical Considerations

Teachers in the research were required to complete consent forms which included ethical safeguards. Teacher participation was voluntary, and teachers were given the right to decline to participate or to discontinue, whether in full or in part, at any time. Parents were also sent consent forms for their children to participate, but these forms were retained by building principals. The school district approved the research design to allow for the collection of data from teachers, students, and trained observers.

Individual teacher data on the self-efficacy scales, the classroom observations, and the self-report inventory were kept confidential by the researcher. Individual reports were provided to each teacher with a summation of the scores from their own observations and their mean student pre- and post-test performance results. The study has only reported group comparison data. Student data were also treated confidentially by project staff and were reported in the aggregate across all the classrooms.



## CHAPTER IV

### Findings

#### Introduction

This study was carried out during the implementation of the pilot phase of Project Phoenix in seven schools in the Norfolk school district. Project Phoenix was a federally funded initiative to develop and disseminate curriculum units designed for use with high ability learners yet employed in the regular classroom with all learners. The initial set of curriculum materials were developed for use in grades two, four, and seven by the Center for Gifted Education at the College of William and Mary. During January, 1999 teachers who had volunteered to implement the units participated in a two day staff development session which focused on the instructional models which were to form the pedagogical infrastructure of the full-blown units. At that time, they were also given 10 - 12 pilot lessons containing the models and asked to incorporate them into their teaching schedule for the second semester of the school year. Teachers who had already covered content embedded in the pilot lessons were asked to use the strategies with different content. The focus of the first wave of implementation was clearly on teachers' capacity to implement these instructional strategies.

This chapter describes the findings from the research that was carried out with this project. The first section of this chapter describes in detail the sample of teachers who volunteered for either the experimental or comparison groups. The second section explains the factor analytical work that was involved in the computation and correlation of scales for the teacher efficacy construct being investigated. The third section focuses on each major research question and sub-questions where applicable and the resultant findings for same. The fourth section summarizes the findings across the research questions.

### Sample

The original sample was comprised of 45 elementary and middle school teachers and their students. The teachers were drawn from seven school buildings in Norfolk Public Schools, most of which had at least 60% of the population on free or reduced-lunch status. Teachers in the experimental group were volunteers. Comparison teachers were drawn from six of the same schools. The size of the experimental group was reduced from 28 to 25 due to attrition. One teacher dropped out of the project citing workload demands, another took maternity leave during the majority of the semester, and a third had a student teacher in the classroom, making it impossible to secure classroom observations. The 17 teachers in the comparison group remained intact. This resulted in 42 teachers being included in the final sample for the study.

### Grade Level

There were 22 second grade teachers, 16 fourth grade teachers, and 4 seventh grade teachers. See Table 2 for a complete breakdown by group.

Table 2

### Grade Level by Group

Gr. Level	Exp.		Comp.		Total	
	Number	Percent	Number	Percent	Number	Percent
2nd	12	48	10	59	22	52
4th	10	40	6	35	16	38
7th	3	12	1	6	4	10
Total	25	100	17	100	42	100

Gender

The sample was composed of 36 women and 6 men. See Table 3.

Table 3

Gender by Group

Gender	Exp.		Comp.		Total	
	Number	Percent	Number	Percent	Number	Percent
Female	22	88	14	82	36	86
Male	3	12	3	18	6	14
Total	25	100	17	100	42	100

Age

The mean age of the sample was 37.7 (SD = 9.1) based on 39 respondents to this item.

The mean age of the experimental group was 38.2 (SD = 8.7) and the comparison group was 37.1 (SD = 10.0). The sample ranged from 24 to 53 years of age. See Table 4 for a complete breakdown by group.

Table 4

Age by Group

Age	Exp.		Comp.		Total	
	Number	Percent	Number	Percent	Number	Percent
29 or less	7	30	4	25	11	28
30 - 39	5	22	7	44	12	31
40 - 49	9	39	2	12	11	28
50 or more	2	9	3	19	5	13
Total	23	100	16	100	39	100

### Level of Education

As shown in Table 5, 69% of the teachers had Bachelor Degrees, and 31% had Masters. About 28% of the sample had no graduate course-work beyond their degree, and 33% had five or more courses. The remaining 39% had between one and four courses. Only 10% of the sample had graduate courses in gifted education; 17% had at least one graduate course in social studies education. One teacher (2% of the sample) had an endorsement in gifted education, and four teachers (10%) were working on one.

Table 5

### Level of Education by Group

Education Factors	Exp.		Comp.		Total	
	No.	Pct.	No.	Pct.	No.	Pct.
<b>BA/BS Degree</b>	15	60	14	82	29	69
0 additional grad. courses	1	7	3	21	4	14
1 - 4 additional grad. courses	7	47	7	21	14	48
5 or more add. grad. courses	6	40	4	29	10	34
<b>MA/MS Degree</b>	10	40	3	18	13	31
0 additional grad. courses	4	40	3	100	7	54
1 - 4 additional grad. courses	2	20	0	0	2	15
5 or more add. grad. courses	4	40	0	0	4	31
<b>Endorsement in Gifted</b>						
Completed	1	4	0	0	1	2
Working On	2	8	2	12	4	10
<b>Grad. course-work in gifted - YES</b>	4	16	0	0	4	10
<b>Grad. course-work in social studies - YES</b>	6	24	1	6	7	17

### Level of Experience

About a quarter of the teachers had two years or less of teaching experience, but 43% had ten or more years of teaching experience as shown in Table 6. Thirty-six percent were new to the Norfolk district, having taught there two years or less, and 69% had been hired by the district within the last five years. Sixty percent of the sample had taught at their respective grade level for at least 3 years.

Table 6

### Level of Experience by Group

Experience Factors	Exp.		Comp.		Total	
	No.	Pct.	No.	Pct.	No.	Pct.
<b>Years of Teaching Experience</b>						
2 years or less	7	28	4	24	11	26
3 - 9 years	6	24	7	41	13	31
10 or more years	12	40	6	35	18	43
<b>Years of Teaching in Norfolk</b>						
2 years or less	9	36	6	35	15	36
3 - 9 years	9	36	7	41	16	38
10 or more years	7	28	4	24	11	26
<b>Years of Teaching at Grade Level</b>						
2 years or less	10	40	7	41	17	40
3 - 5 years	6	24	4	24	10	25
6 or more years	9	36	6	35	15	36
<b>Hire Date in Norfolk</b>						
Since 93 - 94 School Year	15	60	10	59	25	60
92 - 93 School Year or Before	10	40	7	41	17	40

### Comparison of Experimental and Comparison Groups in the Sample

Several analyses were run in order to determine if there were significant differences between the experimental and comparison teachers on key demographic variables. A chi square analysis (Pearson  $\chi^2$ ) was used for the variables of gender, educational degree, and years of teaching experience. A  $t$  test was used for the variable of age. With an alpha set at .05, no significant differences were found between the groups on any of these analyses. (See Table 7.)

Table 7

### Statistical Analysis of Four Sample Variables

Variable	Test Statistic	sig.
Gender	$\chi^2 (1, N = 42) = .264$	.608
Age	$t (37) = .35$	.729
Educational Degree	$\chi^2 (1, N = 42) = 2.366$	.124
Years of Teaching Experience	$\chi^2 (2, N = 42) = 1.423$	.491

### Factor Analysis Component of the Study

Two efficacy instruments were used in this study. One, a measure of global teacher efficacy called the Teacher Efficacy Scale - Short Form (Hoy & Woolfolk, 1990), was adopted intact, and the other, a measure of subject-specific teacher efficacy, was adapted from the Science Teaching Efficacy Beliefs Instrument (STEBI) (Riggs & Enochs, 1991). These instruments were administered to all 45 initial volunteers for the study at the beginning of the first staff development workshop in January, 1999. (See Appendix for copies of the instruments.)

In order to determine if the factor loadings on these instruments were congruent with prior research findings, a factor analysis was conducted for each instrument. The analyses used maximum likelihood extraction with varimax rotation to identify factors with eigenvalues of 1 or

greater. The results for the Teacher Efficacy Scale - Short Form indicated retention of three factors which accounted for 72% of the total variance. Table 8 shows the factor loadings across this study and two prior studies in the literature and presents the means and standard deviations calculated on the basis of the different factor loadings.

Table 8

**Factor Loadings, Means, and Standard Deviations for the Teacher Efficacy Scale- Short Form\***

ITEM	Study Results**			Hoy & Woolfolk Res. ( N = 182 pre-service)		Gibson & Dembo Res. (N = 208 teachers)	
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 1	Factor 2
7	.87			.64		.49	
9	.87			.56		NA	NA
8	.76			.51		.48	
10	.73				.62	NA	NA
6	.67			.60		.51	
2		.86			.59		.60
4		.70			.71		.65
1		.60			.54		.54
3			.61	.48		.53	
5			-.45		.46		.52
Study $\bar{x}$	4.71	3.99	3.77	4.77	3.85		
Study sd	1.10	1.28	0.91	1.07	1.05		
H&W $\bar{x}$				4.2	3.6		
H&W sd				0.6	0.7		
G&D $\bar{x}$						NA	NA
G&D sd						NA	NA

\* Factor Loadings based on 45; means and sd's based on reduced sample size of 42.

\*\* Scale maximum = 6.

Note: Loadings shown are based on the heaviest weight with every item assigned to a factor.

Because the factor loadings derived from study data were different from the factor loadings in the research literature, a comparison of the means of each loading alternative was conducted using a paired sample *t-test*. When the means for the first two factors were compared with the means in the literature from the factor loadings in the Hoy and Woolfolk study (1990), the results were not statistically significant for either the first scale ( $t(41) = .87, ns$ ) or the second ( $t(41) = -1.71, ns$ ). Therefore, the Hoy and Woolfolk factors loadings were used for all subsequent correlations and comparisons involving this instrument. The third factor, identified only in this study, was based on only two items. These items operated in reverse of each other, having loaded on two separate scales on the original instrument. This aberration was not pursued further, and this factor was dropped from subsequent analyses.

The results of the factor analysis for the second instrument, the Social Studies Teacher Efficacy Scale, also deviated somewhat from the STEBI results published by Riggs and Enochs (1991). Three factors (in this case based on the selection of factors with eigenvalues of 2.1 or above), rather than two, emerged and are presented in Table 9. These three factors accounted for 41% of the variance. In fact, 8 factors had eigenvalues greater than .92 which would have accounted for 79% of the variance, but the distribution of items across these factors would have been unintelligible. The low percent of variance accounted for in the three main factors of the instrument (41%) means that 59% of the variance was not accounted for with these factors, thus complicating the interpretation of results.



Table 9

**Factor Loadings, Means, and Standard Deviations for Social Studies Teacher Efficacy Scale\***

ITEM	STUDY RESULTS**			Riggs & Enochs (N = 331)	
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2
21	.84			.69	
23	.83			.60	
22	.83			.72	
24	.78			.69	
19	.65			.76	
8	.60			.68	
5	.56			.69	
3	.54			.67	
2	.53			.54	
12	.47			.75	
17	.47			.75	
6	.42			.64	
16	.41				.52
18	.36			.67	
1		.84			.44
11		.76			.43
14		.72			.61
15		.66			.70
4		.34			.53
7		.31			.57
10		.17			.39
13		-.14			.41
20			.87		.35
9			.63		.35
25			.35		.37

Study mean	3.93	3.42	3.87	3.93	3.57
Study sd	0.48	0.50	0.62	0.50	0.41
R&E mean				4.27	4.12
R&E sd				NA	NA
R&E sum				55.89	49.49

\*Factor loadings based on 45; means and sd's based on reduced sample size of 42.

\*\*Scale maximum = 5.

Note: Loadings shown are based on the heaviest weight with every item assigned to a factor.

A paired sample *t* test was run to detect significant differences between the means based on the original factor loadings and the current study's factor loadings. In this case, the first scale did not show statistically significant differences ( $t(41) = .58$ , ns). However, there were significant differences on the second scale ( $t(41) = 4.69$ ,  $p < .001$ ). This led to a decision to use the study's factor loadings for all subsequent analyses. Although factor one was the same, factor two was more complex, subdividing into a third factor in this study sample.

This third factor, emerging from the study data, appeared to focus on students who are difficult to reach. It suggested that teachers' sense of efficacy takes into account a group of students who are intractable, making a distinction between most students, and a residual core that cannot be impacted. It suggests that teachers can have great impact on most students yet have negative outcomes for some students. This will be discussed in more detail in Chapter 5.

#### Relationship Between the Two Efficacy Measures

One of the sub-questions explored in this study was which measure of teacher efficacy was a better predictor of teacher behavior and student learning. Therefore, two instruments were administered to the same sample, with each instrument having two previously validated scales.

The first scale on each instrument was a measure of personal teacher efficacy (PE); the second

scale was a measure of outcome expectancy (OE). The items on the first scale were based on the use of the pronoun “I” in order to personalize the rating. The items on the second scale focused on what teachers in general can do or cannot do, as juxtaposed with the impact of the home environment. The rationale for linking these items to outcome expectancies was that if all or most teachers could do these things, then positive outcomes were more likely to be assured.

Given the parallel nature of item development across the two instruments, one would expect the scales for factors one and two to correlate across the scales. The results showed the reverse. Scale I of the global measure (GSCAL1) correlated positively with Scale II of the subject-specific measure (SSCOM2), and Scale II of the global measure (GSCAL2) correlated positively with Scale I of the subject specific measure (SSCOM1). The third scale of the subject specific measure (SSCOM3), uncovered only in this study, did not correlate with anything.

These results are presented in Table 10.

Table 10

**Pearson Correlations Across Scales of the Two Efficacy Measures (N = 42)**

	Global Measure		Social Studies Specific Measure	
	GSCAL 1	GSCAL 2	SSCOM 1	SSCOM 2
GSCAL 2	.062			
SSCOM 1	.184	.394**		
SSCOM 2	.499**	-.153	.034	
SSCOM 3	-.073	.201	.079	.265

\*\*p < .01

These results were double-checked as the logical conclusion was an error in coding. This included reviewing the original studies to determine if there were transpositions in factor

loadings as well as re-computing the factor scores in this study to see if there were errors in labeling. However, no errors were uncovered. This result was somewhat troubling and suggested that the instruments did not align with the underlying construct. Therefore, correlations which emerged from these instruments and criterion variables were interpreted with caution.

### Question #1 Results

Question #1 raised the question of whether teacher efficacy is correlated with changes in teacher behavior. Teacher behavior was assessed using the Classroom Observation Form (COF). Experimental teachers were observed by trained external observers four times during the study, twice subsequent to the initial staff development session but prior to the curriculum intervention, and twice during curriculum implementation. Following each observation, the teachers themselves were asked to complete the COF, documenting their perceptions of their own performance. Only experimental teachers were involved in the observations and the self-report process. One of the pre-observation visits for one teacher did not occur so the data from the first pre-observation for that teacher and her observer were counted twice in order to keep her in the analysis.

Sums were computed across the pre- and “during-implementation” (subsequently referred to as “post”) observations. Since the instrument had 40 items, the maximum sum possible was 80. Sums were also computed for each of the nine sub-categories of the instrument, with each sub-category having between three and nine items. Five teachers did not submit one set of post-observation data. Therefore, in order to assure consistency between the two groups, the external observer reports for these five were also eliminated, and the post-test totals were based on doubling the results of the first post-test observation. These adjustments almost unilaterally

reduced the post-observation scores for these five individuals, thus erring on the side of caution.

Prior to answering the major dissertation question, the researcher first investigated two underlying questions. These sub-questions were as follows:

- 1) Is there a change in teacher behavior between the pre- and post-implementation observations?
- 2) Do teachers and external observers have congruent perceptions of the presence or absence of the behavior they are observing?

The answers to both questions precede any understanding of the major research question. If there is no change in teacher behavior over time, one cannot measure the relationship between the change and the construct of efficacy. Similarly, if teachers and external observers have different perceptions of the change, then the relationship between teacher behavioral change and teacher efficacy must be measured for each data source. Means and standard deviations for both the pre- and post administrations of the COF and for both sources of data (i.e., teachers themselves and external observers) are presented in Table 11.

Table 11

**Means and Standard Deviations for the Classroom Observation Form (N = 25)**

Sub-Category for COF	Instrument Maximum	Teacher Self-Report				External Observer			
		PRE		POST		PRE		POST	
		$\bar{x}$	sd	$\bar{x}$	sd	$\bar{x}$	sd	$\bar{x}$	sd
Planning	6	5.84	0.37	5.80	0.50	4.64	1.52	4.32	1.44
Expectations	6	5.52	0.82	5.60	0.96	3.84	1.70	4.88	1.42
Accommodations	10	6.56	2.06	6.76	2.35	5.04	1.67	4.92	1.75
Curr. Del. Feat.	6	4.52	1.45	4.96	1.43	2.92	1.73	3.72	1.59
General Strat.	18	11.80	3.71	12.92	2.41	9.68	3.17	9.96	3.13
Critical Thinking	10	5.68	2.93	7.12	2.28	4.32	2.43	5.96	1.93
Problem-Solving	12	5.32	3.22	6.52	2.63	3.12	2.07	4.04	1.49
Metacognition	6	2.96	1.93	3.68	1.97	1.16	1.43	2.24	1.90
Classrm Extens.	6	1.80	1.76	2.28	2.11	0.68	0.80	1.08	1.32
Total COF Score	80	50.00	13.37	55.56	10.89	35.44	10.20	41.04	10.64

Teachers, who had not been trained on the Classroom Observation Form, consistently rated themselves higher than external observers rated them on all nine categories of the instrument. Both groups reported an increase in the behaviors observed between the pre- and post-observation period in all but two categories. In the “Planning” category, both teachers and external observers reported a reduction in the means for this category, but the teacher reduction appeared negligible. In the category dealing with “Accommodations to Differences,” the teachers’ numbers climbed, but the external observer numbers declined.

In order to answer the two sub-questions, the researcher used a repeated measures multivariate design with both time and rater treated as independent variables. Time examined the

change between the two administrations of the COF; rater examined differences between teachers' self report and external observers' observations. The analysis was run separately for the sub-category scores and the total scores. Mauchly's Test of Sphericity was used to check for equality of covariances and demonstrated a Greenhouse-Geisser epsilon of 1. The results of these analyses are presented in Table 12.

Table 12

Time and Rater ANCOVA for the Classroom Observation Form (COF) (N = 25)

Sub-category of COF	df	TIME		RATER		T X R Interaction	
		F	Sig.	F	Sig.	F	Sig.
Planning	24	0.52	.479	32.71*	.000	0.35	.558
Expectations	24	9.33*	.005	19.20*	.000	6.83*	.015
Accommodations	24	0.01	.925	18.93*	.000	0.36	.553
Curr. Del. Feat.	24	4.26*	.050	15.25*	.001	0.70	.412
General Strat.	24	1.16	.217	25.56*	.000	0.96	.337
Critical Thinking	24	10.06*	.004	7.16*	.013	0.08	.787
Problem-Solving	24	6.18*	.020	20.49*	.000	0.12	.729
Metacognition	24	7.34*	.012	18.83*	.000	0.66	.424
Classrm. Extens.	24	3.04	.094	15.13*	.001	0.06	.802
Total COF Score	24	8.47*	.008	33.97*	.000	.000	.989

This analysis showed that for the total COF score, there were significant differences across both time and rater, and there was no interaction effect. The strength of treatment effect ( $\eta^2$ ) for time was .261 and for rater was .586, both large effects (Cohen, 1988). Across the different subcategories of the instrument, the results were mixed. Four subcategories of the instrument showed significant differences across time for both groups, and these categories all dealt with

behaviors that were specifically incorporated into the curriculum models. These categories and their strength of treatment effects ( $\eta^2$ ) were “Curriculum Delivery Features” (.151), “Critical Thinking” (.295), “Problem Solving” (.205), and “Metacognition” (.235). These strength of treatment effects were also large (Cohen, 1988). All nine categories showed significant differences between raters. Only one category, dealing with “Expectations,” showed an interaction effect. An examination of the means showed that the external observers reported a statistically significant change in this category, but the teachers’ self-reports did not evidence change.

These results suggested that changes in teacher behavior occurred between the pre- and post-observations and that teachers and observers had different perceptions of the experience they were recording. Therefore, correlations between teacher efficacy and changes in teacher behavior were examined separately for each reporter group, i.e., teachers and external observers.

Consequently, correlations were run between the five efficacy sub-scales and the total post observation scores for teachers’ self reports separately from those for external observers’ scores. A hierarchical regression was used, with the pre-observation scores for each group entered into the modeling process first in order to adjust for *changes* in teacher behavior. The results of this analysis are presented in Table 13.



Table 13

**Regression Analysis of Teacher Efficacy Scales (N = 25)**

Model	Predictor Variables	Post Observation			
		df	R <sup>2</sup>	t	sig.
<u>Teacher Self Report</u>					
1	Pre-Observation (Tea. Self Rep)	23	.284	3.02*	.006
(Global) Teacher Efficacy Scale - I					
2	Pre-Observation	22	.343	2.87	.009
	GSCAL1			1.41	.172
(Global) Teacher Efficacy Scale - II					
2	Pre-Observation	22	.341	3.09	.005
	GSCAL2			-1.39	.180
Soc. Studies Teacher Efficacy Scale - I					
2	Pre-Observation	22	.400	3.75	.001
	SSCOM1			-2.06	.051
Soc. Studies Teacher Efficacy Scale - II					
	Pre-Observation	22	.328	2.99	.007
	SSCOM2			1.20	.242
Soc. Studies Teacher Efficacy Scale - III					
2	Pre-Observation	22	.304	3.03	.006
	SSCOM3			0.79	.438

**External Observer Report**

Model	Predictor Variables	df	Post Observations		
			R <sup>2</sup>	t	sig.
1	Pre-Observation (Ext. Obs)	23	.110	1.68	.106
<b>(Global) Teacher Efficacy Scale - I</b>					
2	Pre-Observation	22	.215	1.22	.237
	GSCAL1			1.72	.100
<b>(Global) Teacher Efficacy Scale - II</b>					
2	Pre-Observation	22	.115	1.68	.106
	GSCAL2			0.34	.735
<b>Soc. Studies Teacher Efficacy Scale - I</b>					
2	Pre-Observation	22	.134	1.71	.101
	SSCOM1			0.78	.444
<b>Soc. Studies Teacher Efficacy Scale - II</b>					
2	Pre-Observation	22	.222	1.04	.311
	SSCOM2			1.78	.089
<b>Soc. Studies Teacher Efficacy Scale - III</b>					
2	Pre-Observation	22	.111	1.66	.111
	SSCOM3			0.20	.848

Only the teachers' self-report pre-observations predicted their own post-observations, accounting for about 28% of the variance. There was no relationship demonstrated between the external observers' pre- and post-observation scores, nor between any of the efficacy scales and the post-observation scores, once the pre-observation correlation was removed from the equation.

**Additional hierarchical regressions were run for each sub-category on the Classroom**

Observation Form and the five efficacy scales. In separate analyses, the dependent or target variable was either the teachers' self report post-observation scores or the external observer post-observation scores. The relevant pre-observation score was entered first to account for *changes* in score; the efficacy scale scores entered second. The results detected significant relationships between efficacy scales and seven sub-categories. Three of the sub-categories involved external observer scores; two involved teacher self-reports, and one category involved both. The abbreviated results of only the statistically significant correlations are provided in Table 14.

Table 14

Statistically Significant Regression Correlations Between Efficacy Scales and Sub-Categories of the Classroom Observation Form Post-Observations (N = 25)

Sub-Category	Predictor Variables	Post Observation			
		df	R <sup>2</sup>	t	sig.
Teacher Self Report					
Planning	Pre-Planning	22	.032	0.16	.870
	GSCAL1		.622	5.86*	.000
	Pre-Planning	22	.032	-0.66	.516
	SSCOM2		.413	3.78*	.001
Expectations	Pre-Expectations	22	.564	6.25*	.000
	SSCOM2		.650	2.32*	.030
General Strat.	Pre-General Strat.	22	.242	3.28*	.003
	SSCOM1		.377	-2.18*	.040

External Observers Report					
Planning	Pre-Planning	22	.124	2.60*	.017
	SSCOM3		.395	-3.14*	.005
Curr. Del. Feat.	Pre-Curr. Del. Features	22	.095	1.55	.135
	GSCAL1		.246	2.10*	.048
Critical Think.	Pre-Critical Thinking	22	.016	0.46	.649
	GSCAL1		.199	2.24*	.035
	Pre-Critical Thinking	22	.016	-0.22	.831
	SSCOM2		.241	2.55*	.018
Metacognition	Pre-Metacognition	22	.060	1.01	.325
	SSCOM2		.267	2.49*	.021

In two of the sub-categories of the Classroom Observation Form, “Planning” for the teacher reports and “Critical Thinking” for the external observer reports, two efficacy scales significantly correlated with post-observations. However, these were two scales that also correlated with each other. When both scales were entered into the equation in a subsequent analysis, the variable entered third did not achieve statistical significance. Therefore, they appeared to be duplicating, not expanding the amount of variance explained. Interestingly, these were also the scales that emerged most frequently in detecting correlations, accounting for 5 of the 7 sub-categories in which significant correlations were shown.

Although the teachers’ self reports for the “Planning” sub-category did not change over time, it was curious that only their efficacy scores correlated with their post-test scores. The correlation for teacher self-reports for this sub-category was the strongest correlation in this analysis, with GSCAL1 accounting for 62% of the explained variance. This meant that teachers’

efficacy scores on the personal efficacy dimension of the global measure correlated with their post-observation planning behavior on the self-reports. The issue regarding lack of change also applied to the correlation for the sub-category on “Expectations.”

Two correlations with other efficacy scales, one involving “General Strategies” for teacher self-reports with the social studies specific-measure, Scale I and the other involving “Planning” for external observer reports with the social studies specific-measure, Scale III, were negative, suggesting an inverse relationship. The isolated occurrence of correlations with these measures makes them difficult to interpret.

The correlations between three sub-categories of the COF, namely “Curriculum Delivery Features,” “Critical Thinking,” and Metacognition” with either the first scale of the global measure, the second scale of the subject-specific measure, or both was somewhat interesting. However, most of the correlations were very mild, and again, the lack of consistency across sub-categories of behavior prevented a clear pattern from emerging. In all, 90 correlations were run to test for correlations with sub-categories so perhaps some of these effects were random.

### Summary of Question #1 Results

Teachers and external observers have different perceptions of teacher behavior as measured by the Classroom Observation Form, with teachers indicating a higher number of behaviors in evidence than external observers on the total form and on each sub-category of the form. Both teachers and external observers reported an increase in behaviors observed between the pre and “during implementation” observation cycles on the total instrument and on four important sub-categories related to the curriculum innovation. These categories were “Curriculum Delivery Features,” “Critical Thinking,” “Problem Solving, and “Metacognition.”

The strength of treatment effects for these differences were large (Cohen, 1988).

There were no correlations between any of the efficacy scales and the total post-observations for either group, teacher self-reports or external observer reports (Table 13). Sporadic correlations were demonstrated between some of the efficacy scales and sub-categories of the COF instrument for each group of reporters. The first scale of the global measure and the second scale of the subject-specific measure, which shared common variance, were the more frequently occurring predictors. However, many of these correlations were quite mild, and no clear pattern was able to be detected.

### Question #2 Results

The second research question of interest focused on exploring the relationship between teacher efficacy and student learning. There were three measures of student learning used in the study. One instrument dealt with Concept Learning (CL); another with Critical Thinking (CT), and the third with a Criterion-Referenced Test for Social Studies (CRT) content tied to the school district's Standards of Learning (SOL). These tests were administered on a pre and post basis to both the experimental and comparison classes. Pre-tests for the CL and CT were carried out in February and early March; pre-tests for the CRT started as early as January in some schools, and all post-tests occurred in late May and early June, 1999. The CRT's were administered at the building level by local staff. The CL's and CT's were administered by project staff or central office staff from the gifted program.

The first two measures, CL and CT, were developed by a researcher contracted to Project Phoenix. The first version of these instruments was piloted with 4 - 6 students at each grade level in the district, drawn from the gifted program but not from the pilot schools. The test was

modified based on feedback obtained during the pilot, with the modifications dealing mostly with clarifying instructions by giving examples and streamlining the number of items.

The items on the Concept Learning instrument were related to understanding the meaning of a concept and how it is structured and ordered. Most of the items were repeated from one grade level to the next with additional and more complex items added at subsequent grade levels to increase the difficulty level. The scoring rubric was also adjusted at the seventh grade level so that on items of less difficulty the point values were decreased. The instrument had a maximum of 27 points at the second grade level, 35 points at the fourth grade level, and 27 points at the seventh grade level. Students at the second and fourth grade levels were given a maximum of 25 minutes to complete the instrument. Students at seventh grade had to complete both instruments (CL and CT) in forty minutes or less.

The Critical Thinking measure followed the same path of development. Items on this measure were designed to correspond to Paul's work on critical thinking (1992). The instrument had a maximum of 27 points at the second grade level, 35 points at the fourth grade level, and 32 points at the seventh grade level. Students at the second and fourth grade levels were given a maximum of 20 minutes to complete the instrument.

The CRT was developed by the Research division in the Norfolk district to use in assessing progress on the district's standards of learning. This was the first year the test was in use, and it did not have reliability or validity established. For each objective measured by the test, there were three items. For a student to demonstrate mastery, he or she had to get two of the three items correct. The second grade test had 12 objectives (36 items); the fourth grade test had 14 objectives (42 items), and the seventh grade test had 19 objectives (57 items). Percentile scores

based on objectives mastered were reported by the district for use in the analysis.

The first sub-question of interest for this major research question was to what extent were these three measures correlated. Table 15 presents the results from a Pearson correlation of the pre-test scores for the three measures based on classroom means.

Table 15

Pearson Correlations Between the Three Student Learning Pre-Test Measures

	Pre-Test	
	Concept Learning (N = 42)	Critical Thinking (N = 42)
Critical Thinking	.614*	
CRT in Social Studies (N = 34)	.694*	.761*

\*  $p < .01$

This analysis showed that the three measures were correlated but not highly enough to suggest that they were measuring one construct. However, the extent of the correlations ( $r = .61$  to  $.76$ ) were sufficient to suggest controlling for them through a hierarchical regression approach. The pre-test score for each variable was entered into the equation first, followed by the variable which had the highest correlation with it, then the second highest, then the efficacy measure. In this way the pre-test and inter-test correlations were factored out before testing for any correlations with the efficacy measures. Thus, the major research question was reshaped to investigate whether efficacy scores strengthened the predictive power of any pre-existing correlations. The results of this regression modeling are presented in Table 16.



Table 16

Regression Analysis of Student Learning Measures with Efficacy Scales (N = 34)

Model	Predictor Variables	df	Student Learning Post-Test Scores								
			Concept Learning			Critical Thinking			CRT in Social Studies		
			R <sup>2</sup>	t	sig.	R <sup>2</sup>	t	sig.	R <sup>2</sup>	t	sig.
1	Pre-test	32	.669	8.04*	.000	.463	5.24*	.000	.372	4.36*	.000
Teacher	Pre-test	29	.689	5.92*	.000	.597	3.38*	.002	.511	3.59*	.001
Efficacy	Highest correlated test			0.47	.639		-1.16	.255		-2.52*	.018
Scale - I	2 <sup>nd</sup> Highest correlated test			-1.22	.231		2.72*	.011		1.31	.202
(Model 4)	GSCAL1			0.33	.747		2.03	.051		-0.95	.350
Teacher	Pre-test	29	.688	6.00*	.000	.542	2.87*	.008	.497	3.20*	.003
Efficacy	Highest correlated test			0.60	.553		-0.59	.563		-2.33*	.027
Scale - II	2 <sup>nd</sup> Highest correlated test			-1.30	.203		2.20*	.036		1.52	.138
(Model 4)	GSCAL2			-0.26	.795		-0.40	.690		0.25	.802

Model	Predictor Variables	df	Student Learning Post-Test Scores								
			Concept Learning			Critical Thinking			CRT in Social Studies		
			R <sup>2</sup>	t	sig.	R <sup>2</sup>	t	sig.	R <sup>2</sup>	t	sig.
Social Studies	Pre-test	29	.728	5.81*	.000	.549	2.95*	.006	.501	3.45*	.002
Scale - I	Highest correlated test			0.70	.490		-0.78	.441		-2.38*	.024
(Model 4)	2 <sup>nd</sup> Highest correlated test			-1.35	.187		2.32*	.028		1.39	.176
	SSCOM1			-2.07*	.047		0.76	.456		-0.52	.606
Social Studies	Pre-test	29	.688	5.95*	.000	.575	3.39*	.002	.496	3.38*	.002
Scale - II	Highest correlated test			0.56	.583		-1.03	.313		-2.27*	.031
(Model 4)	2 <sup>nd</sup> Highest correlated test			-1.24	.225		2.40*	.023		1.52	.140
	SSCOM2			-0.10	.932		1.56	.129		-0.10	.921
Social Studies	Pre-test	29	.693	5.91*	.000	.647	3.23*	.003	.496	3.41*	.002
Scale - III	Highest correlated test			0.46	.648		-1.20	.242		-2.37*	.025
(Model 4)	2 <sup>nd</sup> Highest correlated test			-1.32	.198	(.540)	3.36*	.002		1.39	.177
	SSCOM3			-0.70	.488		-2.97*	.006		0.15	.880

The results of this analysis showed that the pre-test for each measure was the strongest predictor of post-test performance, accounting for between 37% and 67% of the explained variance depending on the instrument. However, several statistically significant correlations with other predictor variables also emerged, including some from the scales of the teacher efficacy instruments. The Personal Efficacy Scale (I) of the Social Studies Teacher Efficacy instrument was negatively correlated with the post-test of the Concept Learning measure but added very little to the variance accounted for in the overall model (an increase of 3%). The third scale of this instrument (the one which emerged from the study's factor loadings), showed a mild negative correlation with the Critical Thinking post-test and increased the explained variance from .54 to an  $R^2$  of .65, or 11%. This meant that the lower the scores on the Social Studies Teacher Efficacy - Scale III, or the less that teachers believed that some students could not be impacted, the better the students performed on the Critical Thinking post-test.

In addition, several other correlations emerged which were statistically significant in predicting post-test performance. The Concept Learning pre-test scores added to the prediction equation for the Critical Thinking post-test scores, and the Critical Thinking pre-test scores augmented the predictability of the Criterion Referenced Test in Social Studies post-test scores. The extent of these contributions ranges from 8 - 9%, again quite mild.

Another sub-question of interest emerged related to the availability of data used in several of the previous analyses. This sub-question asked whether there were any correlations between teacher behavior during the observation periods and changes in student learning. To answer this question, a hierarchical regression method was used. Pre-test scores on the relevant measures were entered into the model first in order to adjust for changes in learning. Then the

Classroom Observation Form scores from the pre-observation period were entered. This analysis was run separately for the two data sets, first for teachers' self report scores on the pre-observation, then for the external observers' scores on the pre-observation. The method was repeated with the post-observation scores for each group. In all, 12 regression equations were run.

Only one statistically significant correlation was detected through this procedure, and it was positioned on the threshold at  $\alpha = .05$ . This was a negative correlation between teachers' self reports during the pre-observation period and student scores on the pre-test of the CRT ( $t(18) = -2.120$ ). It meant that the higher the teachers rated themselves during the pre-observation period, the worse their students performed on the pre-test of the Criterion Referenced Test in Social Studies. Because no other correlations were demonstrated in spite of the common variance established among the instruments, this finding was somewhat aberrant.

### Summary of Question #2 Results

Although two correlations were demonstrated between one of two of the scales of the subject-specific efficacy instrument and either the Concept Learning measure or the Critical Thinking measure, these correlations were quite mild and idiosyncratic. The strongest of the two was with the third scale of the Social Studies Teaching Efficacy instrument, the scale that emerged through the factor analysis of the sample's responses. This correlation suggested that the more teachers believed that some students were unreachable, the worse their students performed on the post-test of the Critical Thinking measure. An additional analysis that investigated whether there were relationships between teacher behaviors and student performance based on scores on the Classroom Observation Form and the three student learning

measures uncovered little. A single, borderline, negative correlation was detected between the teachers pre-observation self reports and student performance on the pre-CRT. Again, this anomaly was difficult to interpret.

### Question #3 Results

The third major research question investigated the impact of the curriculum innovation on student learning. This called for a quasi-experimental design as the teachers were not randomly assigned to groups. Student learning was assessed using the three measures described above. Table 17 presents the means and standard deviations for classrooms that had pre- and post-test scores on all three measures.

Table 17

#### Means and Standard Deviations of Student Learning Measures by Group (N = 34)

Measure	Group	Actual Pre		Actual Post	
		$\bar{x}$	sd	$\bar{x}$	sd
Concept Learn.	Exp.	10.04	2.32	12.47	2.03
(Test Maximum = 35)	Comp.	7.97	2.13	10.93	2.59
Critical Think.	Exp.	8.21	2.25	8.57	2.03
(Test Maximum = 35)	Comp.	6.68	2.16	8.02	2.64
CRT in SS	Exp.	51.78	9.83	65.66	12.85
(Scores in Percentiles)	Comp.	46.06	14.71	66.86	17.03

The three measures of student learning were analyzed using a multivariate Repeated Measures Analysis of Variance (ANOVA) to examine time effects and interactions. Mauchly's Test of Sphericity was used to check for equality of co-variances and demonstrated a

Greenhouse-Geisser epsilon of 1.0. Box's Test of Equality of Covariance Matrices and Levene's Test of Equality of Error Variances were also run to check for violations of the corresponding assumptions. These tests came out as non-significant supporting appropriate use of the procedure. The results of the ANOVA are presented in Table 18.

Table 18

**Repeated Measures Analysis of Variance for Experimental and Comparison Groups (N = 34)**

Source	Measure	df	F	sig.	Eta <sup>2</sup>
<b>Between Subjects</b>					
Group	Concept Learning	1	5.92*	.021	.156
	Critical Thinking	1	2.11	.156	.062
	Criterion Referenced in SS	1	0.29	.596	.009
error		32			
<b>Within Subjects</b>					
Time	Concept Learning	1	115.58*	.000	.783
	Critical Thinking	1	7.40*	.010	.188
	Criterion Referenced in SS	1	73.12*	.000	.696
Time X Grp.	Concept Learning	1	1.13	.296	.034
	Critical Thinking	1	2.49	.125	.072
	Criterion Referenced in SS	1	2.91	.098	.083
error		32			

The results of this analysis showed that there were statistically significant gains for both groups across all three measures ( $p < .01$ ) and that there were no interaction effects. This meant that both groups, i.e., experimental and control, made gains over time, but significant differences between the groups were not detected. The difference in the between-groups portion of this

design for the Concept Learning measure meant that the two groups were different to start with and maintained this difference at the post-test.

A sub-question was explored in this phase of the research which investigated the impact of the curriculum on different groups of learners. Only children in the experimental classes were included in this analysis. Four sub-groups were created. The first group was gifted children as identified by the school district's own criteria for inclusion in their pull-out program. The next three groups were created based on pre-test scores from the Criterion Referenced Test in Social Studies. Since districts largely rely on achievement scores to categorize students, this measure was selected because it most approximates an achievement measure. Given the correlations among all three measures, however, almost any measure could have been selected. The low group had scores that were one standard deviation below the mean ( $< 27.7$ ). The high group had scores that were one standard deviation above the mean ( $> 72.3$ ), and the rest fell in the middle. A total of 439 students were included in the analysis. Descriptive statistics for these groups are presented in Table 19.

Table 19

**Means and Standard Deviations for Low, Middle, High, and Gifted Groups**

Measure	Achieve.		Pre		Post	
	Group	Number	$\bar{x}$	sd	$\bar{x}$	sd
Concept Learning (Test Maximum = 35)	low	67	6.49	3.09	8.07	4.29
	middle	279	9.63	4.39	12.22	4.33
	high	75	11.05	4.49	14.19	4.58
	gifted	18	12.50	5.11	15.67	6.30
Critical Thinking (Test Maximum = 35)	low	67	4.91	3.22	5.13	3.79
	middle	279	7.87	3.86	8.40	4.32
	high	75	10.45	4.33	10.48	5.97
	gifted	18	14.00	5.70	15.39	4.89
Criterion Referenced in SS (Scores in Percentiles)	low	67	18.18	7.65	38.49	20.32
	middle	279	50.65	12.44	61.34	21.06
	high	75	83.73	7.55	82.33	15.09
	gifted	18	66.61	16.16	75.28	19.10

These descriptive statistics showed that rankings across all four groups were consistent for the Concept Learning and Critical Thinking measures, with the low performing group having the lowest scores and the gifted group having the highest scores, and that the direction of change between the pre- and the post-tests was positive. However, on the CRT measure the high group outperformed the gifted group on both the pre- and post-tests, but there was a modest decrease in the scores for this group. For the low, middle, and gifted groups, the CRT scores climbed from



pre- to post-test, consistent with the other two measures.

A multivariate Repeated Measures ANOVA was used to test for differences across time and between achievement groups on the three instruments. Mauchly's Test of Sphericity was not significant, but Box's Test of Equality of Covariance was, signaling a problem with the collinearity of the co-variances, which was related to the disparate sample sizes. The results of the omnibus test are presented in Table 20.

Table 20

**Multivariate Analysis of Variance for Gifted, High, Middle, and Low Groups (N = 439)**

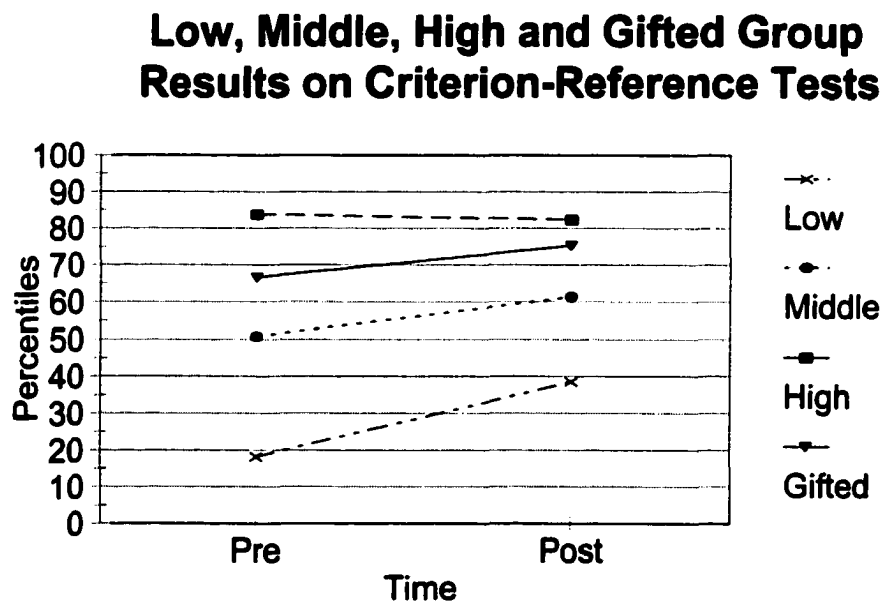
Source	Measure	df	F	sig.	Eta <sup>2</sup>
<b>Between Subjects</b>					
Achieve. Grp	Concept Learning	3	28.94*	.000	.166
	Critical Thinking	3	46.45*	.000	.243
	Criterion Ref. in SS	3	204.69*	.000	.585
error		435			
<b>Within Subjects</b>					
Time	Concept Learning	1	71.84*	.000	.142
	Critical Thinking	1	2.93	.088	.007
	Criterion Ref. in SS	1	49.43*	.000	.102
Time X Ach. Grp.	Concept Learning	3	1.82	.142	.012
	Critical Thinking	3	0.62	.601	.004
	Criterion Ref. in SS	3	16.75*	.000	.104
error		435			

These results showed that there were significant differences on the time factor for the Concept Learning measure ( $p < .001$ ) and no interaction effect. This meant that all the groups

were showing the same rate of change between the pre- and post-tests. Therefore, it was not necessary to run post hoc tests as all groups demonstrated gains.

There were no significant differences over time on the Critical Thinking measure on the omnibus test. Post hoc tests for the Critical Thinking measure were not relevant because there was no indication that any of the groups changed between the pre- and post-test.

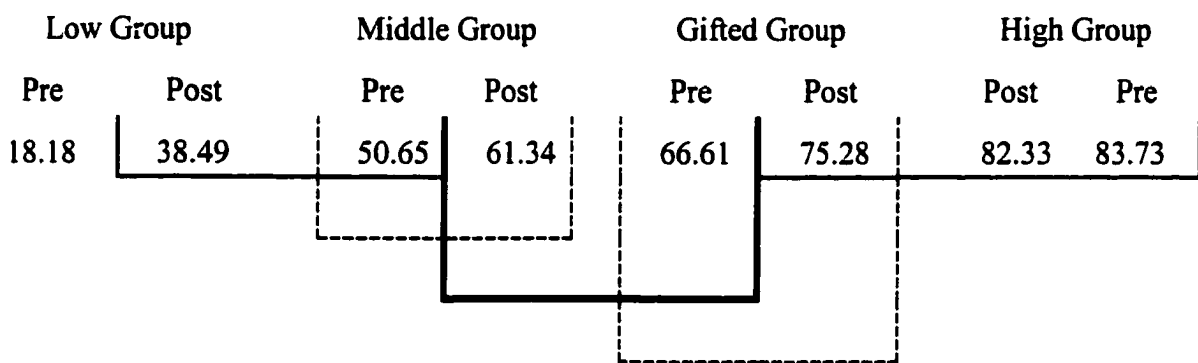
There was an interaction effect on the CRT measure. This meant that changes in performance between the pre- and post-test varied by group membership. In order to display the pre- and post-CRT means, the following graph in Figure 2. was created.



**Figure 1.** Low, middle, high, and gifted group results on criterion-referenced tests.

The question of interest in regard to this interaction was which groups changed significantly and which did not. In order to answer this question, post hoc pairwise comparisons

using the Tukey were run on the pre and post-test means. The critical value that had to be achieved to establish significant difference was 13.80. The results of this post hoc procedure were somewhat complicated, but Figure 2. may help to illustrate what happened. Scores which fall within boxes are not statistically different from each other.



**Figure 2.** Achievement group performance post-hoc test results with open-box connecting scores that are not statistically different from each other.

This figure shows that the low group had significant gains between the pre- and post-tests, and at the pre-test was significantly different from all the other groups. At the post-test, the low group was not significantly different from where the middle group was on the pre-test. The middle, gifted, and high groups did not show significant gains between the pre- and post-CRT as their scores fell within the same boxes. On the post-test, the middle group overlapped with the performance of the gifted group on the pre-test, and the gifted group overlapped with the pre- and post-test performance of the high group. The inter-mingling of the gifted and high group means on the post-test of the gifted group and of the pre- and post-tests of the high group was probably impacted by the ceiling effect. This meant that the true mean for this group of students could not be determined because the test was too easy.

### Summary of Question #3 Results

Although both the experimental and comparison classes showed statistically significant gains over time on all three measures, there were no interaction effects. This meant that how students performed did not vary by the group students were in. An attempt to examine whether or not the curriculum innovation had differential effects on different groups of learners revealed different results for each measure. For the Concept Learning measure, all the groups appeared to make gains ( $p < .01$ ), but for the Critical Thinking measure, none of the groups made statistically significant gains. For the Criterion Referenced Test in Social Studies, there were interaction effects which showed that only the low ability group made significant gains over time ( $p < .05$ ), but two of the groups (the gifted group and the high group) were impacted by the ceiling effect on the measure.

### Question #4 Results

The fourth major research question investigated the impact of the curriculum on teacher efficacy scores. It asked if teacher efficacy scores changed in relation to membership in either the experimental or comparison group. The post-testing on the efficacy scales was done in late May or early June in conjunction with the collection of student learning post-test data. Descriptive statistics for the pre- and post-tests are presented in Table 21.

Table 21

**Means and Standard Deviations for the Experimental and Comparison Groups**

Measure	Group	No.	Actual Pre		Actual Post		$\bar{x}$ diff.
			$\bar{x}$	sd	$\bar{x}$	sd	
Teacher Efficacy Scale - I	Exp.	24	4.70	1.29	4.82	1.14	+0.12
(Scale Maximum = 6)	Comp.	16	4.85	0.72	4.91	0.64	+0.06
Teacher Efficacy Scale - II	Exp.	24	4.14	0.81	4.40	0.92	+0.26
(Scale Maximum = 6)	Comp.	16	3.60	1.19	3.40	1.15	-0.20
Soc. Studies Scale - I	Exp.	24	3.96	0.40	4.02	0.37	+0.06
(Scale Maximum = 5)	Comp.	16	3.96	0.49	3.98	0.32	+0.02
Soc. Studies Scale - II	Exp.	24	3.36	0.55	3.36	0.52	nc
(Scale Maximum = 5)	Comp.	16	3.47	0.45	3.39	0.45	-0.08
Soc. Studies Scale - III	Exp.	24	4.03	0.45	3.90	0.53	-0.13
(Scale Maximum = 5)	Comp.	16	3.67	0.80	3.73	0.53	+0.06

These mean scores showed that on two of the Teacher Efficacy Scales the groups fluctuated slightly in the same direction, and on two others they fluctuated slightly in opposite directions. On Scale II of the Social Studies Teacher Efficacy measure, the experimental group score remained the same, but the comparison group decreased. In order to test for the statistical significance of these changes, the researcher used a multivariate Repeated Measures ANOVA.

In checking for the necessary assumptions, Box's Test of Equality of Covariance Matrices was not significant, nor was Mauchly's Test of Sphericity. However, Levene's Test of Equality of Error Variances was significant for the pre-test of Scale II (OE) of the global

measure, Teacher Efficacy Scale - Short Form. See Table 22 for the results of the omnibus test.

Table 22

Repeated Measures Analysis of Variance for Teacher Efficacy Scales (N = 40)

Source	Measure	df	F	sig.	Eta <sup>2</sup>
<b>Between Subjects</b>					
Group	Teacher Efficacy Scale - I	1	0.20	.659	.005
	Teacher Efficacy Scale - II	1	6.61*	.014	.148
	Soc. Studies Eff. Scale - I	1	0.04	.850	.001
	Soc. Studies Eff. Scale - II	1	0.23	.636	.006
	Soc. Studies Eff. Scale - III	1	2.66	.111	.066
error		38			
<b>Within Subjects</b>					
Time	Teacher Efficacy Scale - I	1	0.22	.642	.006
	Teacher Efficacy Scale - II	1	0.06	.802	.002
	Soc. Studies Eff. Scale - I	1	0.90	.350	.023
	Soc. Studies Eff. Scale - II	1	0.34	.562	.009
	Soc. Studies Eff. Scale - III	1	0.13	.720	.003
Time X Grp.	Teacher Efficacy Scale - I	1	0.02	.888	.001
	Teacher Efficacy Scale - II	1	3.95	.054	.094
	Soc. Studies Eff. Scale - I	1	0.16	.694	.004
	Soc. Studies Eff. Scale - II	1	0.34	.562	.009
	Soc. Studies Eff. Scale - III	1	1.17	.285	.030
error		38			

This analysis showed that there were no significant changes over time within any of the five efficacy scales used in the study, nor were there any interaction effects. However, the analysis did detect a statistically significant difference between the experimental and comparison group on Scale II of the Teacher Efficacy Scale - Short Form, suggesting that the groups were different to start with on this scale, in spite of the fact that neither group changed over time. However, this was the scale that did not meet the assumption for equality of covariances.

#### Summary of Question #4 Results

The curriculum intervention did not appear to impact the efficacy scores of teachers. There may have been a difference in the two groups on their pre-test scores on one of the scales, but it was more likely that a violation in one of the assumptions underlying the use of this statistical procedure triggered the anomaly.

#### Summary of Findings

The research findings for this study were grouped into two categories. The first category dealt with findings related to the construct of teacher efficacy, and the second category focused on evidence of the effectiveness of the curriculum innovation during the pilot phase of Project Phoenix.

Findings related to teacher efficacy:

- 1) Neither the global measure of teacher efficacy nor the subject-specific measure of teacher efficacy performed in the same way as in earlier research findings from the literature. Factor analysis of both instruments uncovered a third factor. On the global measure, this factor seemed unintelligible, but on the subject-specific measure, it appeared to detect a factor related to lack of success with some

students. Factor II on the original measure, the STEBI, focused on outcome expectancies of teachers. The introduction of the third factor on the Social Studies Teacher Efficacy instrument suggested that outcome expectancies could be positive or negative, and this third factor was isolating the potential for negative outcomes with some core group of students.

- 2) Although derived from the same theoretical construct regarding the two dimensions of teacher efficacy (Personal Efficacy and Outcome Expectancy), the corresponding scales on the two measures of teacher efficacy did not correlate with each other but with the alternative factor (i.e., Scale I of the global measure correlated with Scale II of the subject-specific measure and vice versa). These correlations were not strong but were indeed curious. The third scale did not correlate with any of the other scales, again confirming the isolation of a new factor in the complexity of the construct as measured by this instrument.
- 3) Since one of the areas of interest in this study had to do with whether a global measure or a subject-specific measure of teacher efficacy was a better predictor of teacher/student performance, the inverse connection between the two scales of the two instruments confounded the examination of this issue. Additionally, no pattern was discernable from the limited relationships that were detected.
- 4) No relationship was detected between any of the efficacy scales and changes in teacher behavior as measured by the total scores for the Classroom Observation Form. This was true for both teacher self report and external observer data. However, teacher self report on the pre-observation did significantly correlate



with teacher self report on the during-implementation observation.

- 5) Diffuse and inconsistent correlations were detected with some of the sub-categories of the Classroom Observation Form and some of the scales of the composite efficacy instrument.
- 6) A weak negative relationship was detected between the first scale (PE) of the Social Studies Teaching Efficacy instrument and changes in student learning ( $p < .05$ ) on the Concept Learning measure, adding 3% of explained variance. A mild negative relationship emerged between the third scale of the same instrument and changes in student learning on the Critical Thinking measure ( $p < .05$ ), augmenting the explained variance by 11%. The direction of these two statistically significant relationships indicated that the higher the teacher's score on the efficacy scale, the lower his or her students performed on the relevant learning measure. None of the other scales significantly correlated with any of the student learning measures, including the Criterion Referenced Test in Social Studies.
- 7) Pre- and post-scores on teacher efficacy scales did not change during the time period of the study nor were there interaction effects attributable to group membership.

#### **Findings related to curriculum innovation:**

- 1) Both teachers and external observers reported a statistically significant increase in total behaviors observed on the Teacher Observation Form between the pre- and during-implementation observations ( $p < .01$ ) as well as on four categories of the

form ( $p < .05$ ). These four categories, “curriculum delivery features,” “critical thinking,” “problem-solving,” and “metacognition,” were specifically related to the instructional models embedded in the curriculum innovation.

- 2) Teachers self-reported higher levels of specific behaviors than did external observers on all categories of the Classroom Observation Form ( $p < .01$ ).
- 3) All three measures of student learning, the Concept Learning instrument, the Critical Thinking instrument, and the Criterion Referenced Test in Social Studies, were correlated with each other but not highly enough to suggest that they were measuring the same construct.
- 4) There were statistically significant gains ( $p < .01$ ) on all three measures for both the experimental and comparison student groups between the pre- and post-tests, and there were no interaction effects related to group membership in the experimental or comparison classes.
- 5) When achievement groupings of the experimental student group only were examined to see if the curriculum innovation had differential effects on low, middle, high, and gifted students, the Concept Learning measure showed statistically significant gains across all groups ( $p < .001$ ) but no interaction effects. This meant that there were no differential effects detected.
- 6) The Critical Thinking measure showed no gains for any group. The reason that there were gains across time for the total group on the Critical Thinking measure, but not for any of the sub-groups in this additional analysis had to do with the unit of analysis shifting from the class level in the first statistical comparison

(See item 4 above) to the student level in the secondary analysis.

- 7) The results from the Criterion Referenced Test on ability groupings showed an interaction effect ( $p < .001$ ). Post hoc tests for the Criterion Referenced Test in Social Studies revealed that only the low group of students made significant gains from the pre- to the post-test on this measure. Gains for the gifted and high group appeared to be impacted by the ceiling effect of the instrument.

The next chapter discusses these findings in more detail, draws some conclusions regarding them, and suggests implications of the study for further research and practice.

## CHAPTER V

### Discussion, Conclusions, and Implications

This study combined the investigation of two key dimensions: 1) an examination of the construct of teacher efficacy in terms of its relationship to several foci of curriculum reform, and 2) the impact of a curriculum innovation in social studies on student learning. Each of these dimensions are discussed separately in this chapter.

#### Teacher Efficacy and Its Relationship to Teacher Behavior and Student Learning

The literature is replete with examples of the relationship between teacher efficacy and teacher behavior and student learning (Ashton, 1985; Ashton & Webb, 1986; Ross, 1995). Many of these studies have focused on teacher willingness to engage in educational innovation or with teacher classroom behaviors such as persisting longer, providing greater academic focus, and using positive feedback (Rose & Medway, 1981; Gibson & Dembo, 1984). Furthermore, many of these studies have used teacher self-report as the basis for inferring teacher behavior (Ross, 1995). Studies on the relationship between teacher efficacy and students have tended to use standardized achievement tests as the measure of student learning or have examined its relationship to affective characteristics such as student efficacy or motivation (Tschannen-Moran et al., 1998). Many studies have focused on the global measurement of efficacy; a few have targeted it more specifically based on subject matter and grade level considerations (.Pajares, 1996; Tschannen-Moran et al., 1998).

#### Positioning the Study in the Field

This study was designed to extend the literature in several ways. First, the study used multiple data sources to investigate teacher behavior; one was teacher self-report and the other was a report by external observers. This allowed for the relationship between teacher efficacy

and teachers' perceptions of performance to be compared with the relationship between the construct and independent assessments of performance. Thus, comparisons could be drawn between respondent groups on the observation instrument and the construct under study to assess the locus and magnitude of potential correlations.

Secondly, the classroom observation instrument itself was created to incorporate specific behaviors related to educational reform such as setting high standards for the lessons, using strategies that require critical thinking and problem-solving skills on the part of students, and attending to metacognitive experiences. Since no other observation instrument in the field of gifted education was found that targeted classroom teaching behaviors directed to high ability learners within the context of the reform agenda, the researcher, in collaboration with her advisor, developed the Classroom Observation Form. Psychometric data on content validity and inter-rater reliability were calculated for this instrument in order to use it in the study. This gave the researcher a chance to investigate whether the relationships that had been documented previously with good general classroom practice were sustained when the task demands for teacher performance were elevated.

Thirdly, this study used alternative assessments of student learning that were also tied to the reform agenda. Two of these measures, the Concept Learning and the Critical Thinking instruments, were designed to measure higher order thinking skills, and the third measure, the Criterion Reference Test in Social Studies, was designed to measure student performance on the standards of learning that were the basis of the social studies curriculum in the local district. Unfortunately, all three of these instruments were in their first year of development which may have impacted their precision in documenting changes over time and between groups.

The fourth contribution of this study was intended to relate to the measurement of teacher

efficacy. One of the dimensions along which efficacy expectations can vary is the dimension of generality, the extent to which the judgment of efficacy is broadly or narrowly defined. By employing both a global and a subject matter-specific instrument, the study afforded the researcher an opportunity to investigate which approach, if either, was more powerful in predicting behavior.

#### Theoretical construct underlying the study

Teacher efficacy is derived from Bandura's work on self-efficacy which suggests that cognitive beliefs are powerful determinants of behavior. Teacher efficacy, which deals with teachers' perceptions of their ability to impact student learning, has at least two components; personal teacher efficacy relates to teachers' beliefs about their own competence and outcome expectations (general teacher efficacy) relates to how performance mastery impacts student learning outcomes. These two components are related but can operate independently.

#### Discussion related to the measurement of efficacy

Instrumentation to investigate teacher efficacy has evolved from two theoretical traditions, one from Rotter's work on internal versus external locus of control, the other from Bandura's work on self efficacy and its application to the field of education. For this study the researcher selected instruments derived from Bandura's model. One instrument, the Teacher Efficacy Scale - Short Form, was adopted intact as a measure of global teacher efficacy. This scale had originally been constructed by Gibson and Dembo (1984) and further validated by Woolfolk and Hoy (1990) who had reduced it to the 10 items that had loaded most heavily on its two scales. However, they urged researchers "to conduct factor analysis on their own data, because the loadings have not always been consistent across studies" (Hoy & Woolfolk, 1993, p. 213).

Guskey and Passaro (1994) also reported problems with the longer version of the instrument based on the fact that one factor's items were all positively worded, and the other factor's items were all negatively worded, a condition that also applied to the shortened version. However, when they adjusted the items to correct for such contamination, they still found two separate factors that were only moderately correlated. What remained in contention was the meaning of the two factors. Instead of representing efficacy versus outcome expectations, they suggested that they represented the dimensions of personal control versus elements "beyond the control of individual teachers" (Guskey & Passaro, 1994, p. 639).

The results from this study's factor loading were inconsistent with the Woolfolk and Hoy (1990) loadings, but the statistical analysis of the differences showed that they were insignificant. The results did, however, raise questions about the stability of the instrument as a global measure of efficacy. Two of the ten items loaded on a third factor, and this factor was created by extracting one item from each of the other two factors. There was no apparent way to interpret this third factor independently from the other two, given the nature of the items.

Similarly, the factor analysis of the second instrument, the social studies-specific instrument, adapted from one in the literature based on science teaching efficacy, also showed the emergence of three factors, but the third cluster of items was more discernable. In comparing the means of the factor loadings of the first two factors with those in the literature, it was found that the means for the first scale were not statistically different, but the means for the second scale were. Therefore, the second factor in the Riggs and Enochs (1990) study, became two separate factors in this study, both relating to outcome expectancy beliefs. The distinction between the two factors was in regards to positive and negative outcomes. The emergence of this third factor suggested that teachers could have high efficacy beliefs about the effectiveness of the

profession of teaching, but still believe there is a residual core of students that the profession cannot successfully impact.

This finding was not unlike what Guskey (1987) had recognized in terms of the findings from his own efficacy scale, but it was interesting that it emerged on the instrument being used in this study. Is social studies less threatening to teach than science so teachers are more comfortable with attributing potential for failure to the students themselves? There is evidence that elementary teachers feel unprepared to teach science (Riggs and Enochs, 1990), but this same trepidation may or may not extend to social studies. Another possibility is that teachers have become more comfortable with self assessment generally and assessment of the efficacy construct in particular as it has proliferated in the literature and so are willing to make finer discriminations in reporting their beliefs. Clearly awareness of the construct has grown as it has continued to be investigated and reported on in the literature. A third possibility is that in urban settings, with high numbers of economically disadvantaged students, teachers are less optimistic about their abilities to reach the entire group, perhaps feeling that the odds are stacked against them. One can only speculate as to why this distinction occurred in this study.

The inability of the adapted instrument to account for a significant percentage of the variance was also problematic in this study. Since only 41% of the variance was attributable to the three factors, 59% was unaccounted for. Therefore, any correlations that emerged could have been derived from these unidentified factors and not even represent the theoretical dimensions of the construct under study.

The results of this study clearly supported what the literature in general has recently addressed; namely, that the measurement of teacher efficacy is very complex and our instrumentation very weak. Even instruments that had some track record in the literature



performed inconsistently in this study. Tschannen-Moran et al. (1998) have advanced this issue by offering a new conceptual model that stresses the context-specific dimension of efficacy by incorporating an analysis of the teaching task and its context into the measurement instrument. Although their model has yet to be tested, it offers greater promise for a deeper understanding of this important, yet elusive, phenomenon.

#### Discussion Related to the Relationship between Teacher Efficacy and Teacher Behavior

The findings in this study did not uncover a relationship between teacher efficacy and changes in teacher behavior based on total scores, either in terms of teachers' self reports of behavior or external observers' reports of behavior using the Classroom Observation Form. Although there was an increase in total teacher behaviors associated with the instrument from the "pre" to the "during implementation" period, there were no correlations with any of the efficacy scales. There are several explanations for this which might be considered. First, such relationships might not exist. Clearly this study was not able to disprove that assertion. Earlier studies that have found relationships have been fairly narrow in their identification of specific teacher behaviors (Ashton, 1985; Gibson & Dembo, 1984). It could be that as teachers are assessed on a broader array of behaviors that include implementing more sophisticated strategies in the classroom, the relationship with teacher efficacy is not maintained. A second possibility is that the structural problems with the efficacy instruments prevented the detection of such a relationship. These problems have already been addressed previously, but an implication of reliable instrumentation is increased potential for error. In this case, a false negative finding that indicated no relationships existed when, in fact, they really did. Thirdly, the limited sample size may have obscured the detection of a relationship. With only 25 individuals, there may not have been a critical incidence level high enough to find the relationship.

Additional analyses run with the sub-categories on the Classroom Observation Form (or the Teacher Self-Report Inventory) and the five efficacy scales detected some mild to strong relationships. These results suggested that some of the possible explanations cited above might not apply, such as inadequate sample size and weak instrumentation. In fact, what these correlations suggest is that it is easier to detect relationships within categories on the COF (or the Teacher Self-Report Inventory) than across it, speaking again to the issue of narrowness and/or specificity of behavior. Of course, for correlations with the subject-specific scale, the unexplained variance in the instrument may be a consideration. The correlations may be based on the factors that are not accounted for, rather than the ones that are. The biggest issue with the detection of these sub-category correlations is their lack of consistency. No one scale stood out across all the correlations that were established, and most of the correlations were mild. However, in the sub-categories of the instrument that were most related to the teacher behaviors stressed in the curriculum (i.e., expectations, curriculum delivery features, critical thinking, and metacognition) one of two scales were correlated (or both). These were Scale I of the global teacher efficacy instrument, dealing with personal efficacy, and Scale II of the subject-specific teacher efficacy instrument, dealing with outcome expectancy. These two scales were also correlated with each other. This might suggest that as specific clusters of behaviors are identified related to sophisticated pedagogical practices, some dimensions of the construct of efficacy may be correlated. At the global level, it is the personal dimension of the construct, but at the subject-specific level, it is the outcome (or general) dimension of the construct. However, given the inconsistency of the performance of the instrument over the multiple data sources and all the sub-categories, this tentative conclusion should be treated with caution. Consequently, this researcher is reluctant to draw strong inferences from these data.

### Discussion Related to the Relationship between Teacher Efficacy and Student Learning

Most of the previous research investigating the relationship between student learning and teacher efficacy was conducted with global measures of efficacy, and neither of the scales of the global measure used in this study established any correlation. Small but statistically significant correlations were found between two scales of the subject-specific teacher efficacy instrument and two measures of student learning. Scale I of the Social Studies Teacher Efficacy instrument showed a mild negative correlation with the Concept Learning measure, and Scale III showed a mild negative relationship with the Critical Thinking measure. This meant that the higher the teachers' sense of efficacy on these two factors, the lower the students' test scores on the respective instrument.

Previous studies had established positive correlations with measures of student learning, which contributed to the excitement about this construct (Ashton & Webb, 1986; Guskey, 1982), starting with the RAND studies in the seventies (Armor et al., 1976). However, these earlier studies also used traditional measures of student achievement as the proxy for student learning (Tschannen-Moran et al., 1998). The closest measure of achievement in this study was based on the Criterion Referenced Test in Social Studies, and no correlations were established with this instrument. Also, this study focused on changes in student learning as measured by factoring out the pre-test performance in the equation.

Again these findings were difficult to interpret. No clear pattern emerged with any of the scales, and the correlations that were detected were relatively mild, albeit negative. This might suggest that as measures more akin to student aptitude are used to measure student achievement, the traditional patterns are reversed. However, given the psychometric short-comings of the subject-specific teacher efficacy scale (inability of three factors to account for more than 41% of

the variance), it would be safer to conclude that the study results were inconclusive in terms of answering this question.

### **Discussion Related to Curriculum Innovation**

This section of the chapter discusses the findings in relation to research questions dealing with the impact of the curriculum innovation. It is divided into three sub-sections with the first focusing on teacher behavior, the second on student learning, and the third on sub-groups of students based on achievement.

#### **The Curriculum Innovation and Teacher Behavior**

Perhaps the most interesting findings from the study were those related to the introduction of the curriculum innovation into the classroom. One of the promising findings in this regard was that teachers observed using the new social studies curriculum strategies showed more evidence of using behaviors targeted toward higher-order thinking and problem solving skills. These behaviors were in the sub-categories of the Classroom Observation Form dealing with “Curriculum Delivery Features,” “Critical Thinking,” Problem-Solving,” and “Metacognition.” Since an experimental design was not used to test this question, this could be attributable to another variable in the equation such as time. Therefore, one cannot attribute causality to changes in teaching behavior from the pre- to the during-implementation observation period to the curriculum innovation and training that accompanied its utilization.

One can, however, note that both teachers and external observers saw increases in specific categories that were tied to the teacher training experience and curriculum design during the implementation cycle of observations. The directionality of these changes coupled with the alignment of the intervention with the specific categories of the instrument that reflected the growth may suggest evidence of positive impact. Since the two day training was focused on

instructional models addressing these skills and the curriculum itself provided a scaffold that promoted the teaching of higher order thinking skills and problem-solving, one might tentatively consider that these elements impacted teacher performance during the post-observation period. Furthermore, if one assumed that over time or due to maturation that teachers improved in their ability to demonstrate increased numbers of behaviors, one would assume that this would be reflected across the instrument, not in isolated categories. Since there was not a concomitant rise in the numbers of other general teaching strategies during the same observation cycles, one is more likely to consider the possibility that the curriculum innovation made some modest improvements in teachers' performance related to enhancing higher order thinking, problem solving, and metacognition in the classroom.

It was also interesting that teachers gave themselves higher scores on their version of the Classroom Observation Form than external observers. This meant that teachers were more likely to think that they were implementing certain behaviors than external observers concurred with in their observations. There was no way to tell from the data whether teachers were over-estimating their own performance or external observers were under-estimating teacher performance. There was far greater variability in the data reported by external observers, as indicated by the size of the standard deviations. However, given the range of ages, experience, and education evidenced in the sample of teachers, one might also expect a higher level of variability in performance.

If one were to speculate that teachers were over-estimating the incidence of sub-categories of behavior, this would be somewhat troubling. Since the teaching of convergent and divergent thinking and reasoning skills is quite complex and requires abstract thinking, one might suspect that teachers would under-estimate their abilities to implement these behaviors. If this interpretation were accurate, it might mean that teachers do not understand the complexity of

pedagogy that uses these skills.

It will be interesting to see in year two of the project, after the teachers have received training on the instrument, whether their perceptions will be more closely aligned with the external observers. This should provide more data from which to determine which group of reporters, if either, shifts their relative position.

Only one correlation between teacher behavior on the Teacher Self Report Inventory (their version of the Classroom Observation Form) and student learning was detected. This was a negative correlation between teachers' pre observation self reports and student performance on the pre-test of the Criterion Referenced Test in Social Studies, suggesting that the higher teachers rated themselves on the pre observation, the lower their students' performance on the pre-test of the CRT. This might mean that this disconnection between what teachers think they are doing versus the reality of what they are actually doing has negative consequences for how their students perform. However, this correlation was not sustained with post-CRT scores, and if this were a legitimate explanation, there would be no reason that it would disappear at that stage of testing. Given that this correlation was limited to pre-test results only and no other correlations were established for the other measures or between the external observers' reports and the student measures, perhaps one should not ascribe too much meaning to it.

Therefore, the more salient question is why did no correlations emerge between these data sets. If, in fact, teacher behaviors related to higher order thinking instructional strategies increased, why were not these changes correlated to student performance, particularly in regards to measures of Concept Learning and Critical Thinking? The answer to this question may be that the teacher changes were detectable but not sustained over time. Teachers exposed to two day training on selected pedagogical practices and given pilot lessons to implement were able to

show greater evidence of relevant behaviors during the post-observation cycle. However, the depth of these changes may not have been sufficient to impact student learning at this stage of implementation.

### Curriculum Innovation and Student Learning

There was no evidence that the curriculum innovation made a difference in student learning on the three measures employed in the study. The fact that both the experimental and comparison classes made gains on all three instruments at least showed that the innovation was not harmful. Furthermore, the lack of impact of the curriculum innovation was fairly easy to understand as the implementation of the curriculum during the semester appeared to be quite shallow and uneven across sites. Although teachers in the experimental group had agreed to allocate 25 hours of instructional time to the new curriculum, follow-up data obtained through the debriefing process revealed that no teachers taught all the lessons, and many teachers taught only 4 - 6 of the lessons, constituting no more than 6 - 15 hours of instructional time. Data collected at the end of the semester documented the numbers of lessons each teacher taught.

In addition, feedback from external observers indicated that teachers struggled with lesson implementation, in some cases taking a 45 minute lesson and extending it over seven class periods, thereby losing the continuity that was central to the lesson's purpose. Other teachers acknowledged similar experiences during the summer debriefing process. While these adaptations may have addressed the time commitments made by teachers to the project, they did little to improve the quality of instruction in the classroom in a meaningful way. In fact, had significant gains in student performance been detected, they would have been seen as extremely suspicious by project staff knowledgeable about actual implementation.

### Discussion of Student Achievement Sub-group Results

As student sub-group performance was examined, the findings were somewhat curious. The findings showed that all groups grew on the measure of Concept Learning; no groups grew on the measure of Critical Thinking, and only the low group grew on the Criterion Referenced Test in Social Studies. One has to ask why would student groups show changes on a measure of Concept Learning but not on a measure of Critical Thinking? One explanation for this might be related to development or maturation. This might be investigated further by looking at individual grade level gains, but they were not investigated under the scope of this study. Another explanation has to do with the order of testing. These two instruments were given in tandem in the same sequence. It was clear to test administrators that students at the second and fourth grade levels had difficulty sustaining focus through the completion of the testing cycle. The post-testing at several sites occurred the last few days of school, with temperatures in the high 90's and no air conditioning. It would be prudent to consider varying the order of the testing in subsequent years to determine if this issue contributed to the result.

The fact that all achievement groups showed increases on Concept Learning was very illuminating as one might have expected differential learning gains for groups of learners. This finding suggested that the low and middle groups got as much out of exposure to the experience as the high and gifted students did. Since one of the concerns often raised by teachers is that the introduction of higher standards might leave the slower learners in the dust, this finding suggests the opposite. In fact, everyone gained when gains were made. Again, some caution is warranted here. This analysis was only done with the experimental classes so no causality can be attributed to the curriculum innovation, but follow up analyses to examine patterns with the comparison classrooms are clearly warranted at some future point.



The fact that the gifted and high ability students did not progress on the CRT may be partially explained by ceiling effect on the test. Ceiling effect means that the performance of the group regresses toward the mean based on random error in test scores. Since the ceiling of the test limits the top score, the error can only work against the student's performance, not for it. The high ability group had its mean near the top of the instrument on the pre-test and stayed at the same level for the post-test, although the standard deviation almost doubled. The gifted group also did not show significant growth between the pre- and post-test, although their variability increased only moderately. The increases in variability suggested that something was going on with the upper groups, even though overall mean performance was not significantly impacted. It might have been that higher performing students were frustrated with test-taking by the end of the year so some portion of the group did not demonstrate their optimal performance capacity. It is well established that the optimum environment for gifted learners involves both acceleration and grouping (Kulik & Kulik, 1992; VanTassel-Baska, 1992b). Since the social studies intervention in the elementary grades did not involve a grouping component, this may have had a deleterious effect on the small group of identified learners as well as some students who were not formally identified.

Only the low group of students, not the middle group, made gains on the CRT. These gains were accompanied by increases in variability, again speaking to the impact of instruction in differentiating performance within this group. It could be that teachers focused instruction at the threshold of the low students so that was where progress was demonstrated. An examination of the second and fourth grade instruments by the researcher showed that most of the items were for recall of historical facts, although a few items tested application of map reading skills. By teaching the knowledge necessary to do well on this measure, teachers would be likely to

emphasize lower-order performance demands. Therefore, the students who scored the lowest on the pre-test were most likely to improve their scores by learning the new facts. However, students who started at a higher threshold did not have instructional experiences that concomitantly increased their mastery, as teachers may have been covering ground they already knew.

### Conclusion

This research was undertaken with two major foci in mind, one dealing with the construct of teacher efficacy and its relationship to teacher behavior and student learning, the other focusing on the impact of a curriculum innovation on student learning and teacher efficacy. The first component of the study yielded little new data but confirmed the present difficulties in measuring this intriguing, but elusive construct. No clear pattern emerged regarding the relationship between personal or general (outcome) efficacy and teacher behavior or student learning. Whether this was related to measurement problems or to the failure of the construct to correlate with higher task demands on the part of teachers and students is not known. Also, no project impact on teacher efficacy was demonstrated. Further research, deploying more sophisticated instrumentation, will be necessary to secure more definitive answers to these questions.

The second component of the study offered some promising insights about the potential for the new curriculum to change classroom practice. While it would be erroneous to conclude that the staff development and curriculum intervention that were central to this curriculum innovation *caused* changes in teacher behavior, one might be optimistic that they may have contributed to potential changes in behavior related to the teaching of critical thinking, problem solving, metacognition, and certain curriculum delivery features. Since both the external

observers and the teachers themselves reported an increase in these behaviors, there was a consistent perception that such increases occurred. The compatibility of these perceived changes with the nature of the innovation gives some credence to this interpretation of the data. More conservative researchers would require tighter experimental controls before conceding this point. Furthermore, the disparity between teacher self-reports and external observer reports warrants some attention. Although one cannot determine from the facts in evidence which group is out of kilter, it is possible that teachers are over-estimating their performance in the classroom.

Unfortunately, perceived changes in teacher behavior were not deep enough to translate into student learning gains at this stage of implementation that exceeded the performance of the comparison group. As teachers become more sophisticated in their use of the teaching models, have the opportunity for practice effect, and teach a full-blown unit of study involving at least 25 hours of instruction, a better test of the impact of the curriculum on student learning can be undertaken. New instrumentation to measure concept learning, critical thinking, and teacher practice appears promising, but the use of a criterion referenced content-based measure proved ineffectual.

### Implications for Research and Practice

If teacher efficacy is to be pursued as a valuable construct in advancing the educational reform agenda, a first step is clearly to devote more attention to the development of instrumentation that more accurately measures the construct. This study did little to advance our understanding of this construct and how it plays out against increasing demands for teacher and student performance, other than to point out the flaws of the existing methodology. Questions regarding the generality of the measurement of the construct as well as the range of teacher and student behaviors that it may predict remain unanswered. Further researchers would be advised

to consider the template advanced by Tschannen-Moran et al. (1998) in creating an instrument which is more sensitive to the behaviors required in the context, or to the more sophisticated and extensive work of Curda (1998) in tailoring an instrument to the unique dimensions of specific subject matter areas and in changing the approach to measurement of the outcome dimension.

There is also a need to revise and strengthen the psychometric properties of the student learning measures employed in this study. The Criterion Referenced Test in Social Studies has a problem with ceiling effect which needs to be addressed if it is used again by the school system. Clearly it is inappropriate for Project Phoenix work and served as a poor proxy for the content-based dimensions of the curriculum lessons. Should the district decide that it is useful for internal reporting purposes, district staff should conduct an item analysis. Some of the troubling issues with this measure have to do with why the middle group of students did not show gains over time and why were there such large increases in variability with several of the sub-groups.

The two new instruments for Concept Learning and Critical Thinking should undergo similar item analyses. If the length of these instruments could be curtailed, students might be able to sustain their interest for longer periods, thus increasing the reliability of the measures. While the instruments appeared to perform reasonably well for the pilot phase, their psychometric properties could be improved over time.

In regard to the impact of the curriculum innovation, this researcher would make several observations. First, it is important to recognize that teacher behavioral change is a complex process that takes time. Data that are gathered in subsequent years of the project may be more likely to illuminate the capacity of the curriculum to increase student reasoning and problem solving behavior, once teachers have reached a better understanding of the instructional models and have had practice effect with them. The limitations of the research design itself, the short-

time frame for the implementation of this pilot, and the limited application of the instructional models as reported by the teachers themselves, were insufficient to achieve deep changes in teacher behavior that resulted in student learning gains attributable to the innovation. Evidence of nascent changes reported by both teachers themselves and external observers, however, was quite promising. Fullen (1991) has done much to describe the complexity of change in the educational arena, but time continues to be the enemy. By the conclusion of the first year of the project, at least six experimental teachers had left the system or moved to different positions. So, even as the involvement of the teachers increases, attrition takes a concomitant toll.

A second observation for practice is to support teachers in the acquisition of new behaviors by increasing opportunities for dialogue between the external observers and the teachers about what each is seeing in the classroom. Because there was no opportunity for sharing perceptions during the pilot phase, teachers' essential understandings did little to change. Since they thought they were doing what they should have been doing, it came as a surprise to them that the external observers saw it differently. Disclosing this at the end appeared threatening rather than helpful to the teachers, with several of them remarking that they had not been trained on the instrument.

Third, there is a need for teachers to capitalize on other support strategies such as peer coaching and demonstration teaching. The fact that the project was in seven different schools and staff assigned to assist with the project carried other full-time responsibilities, limited the availability of other supports that have been shown to be effective. Although demonstration teaching was offered to project teachers, very few took advantage of it. One of the reasons for this may have been a perceived rivalry between the schools and the central office gifted program staff. If teachers were to request a demonstration of a lesson, they may have felt they were

admitting that they could not do it themselves. Nevertheless, it would be important to ensure that more teacher supports are in place so that teachers acquire a deeper understanding of and the capacity to apply the instructional models that underlie the curriculum lessons.

### Summary

Standards-based educational reform has created new challenges for both educators and students alike. This study was undertaken to explore how some of the pieces of the reform agenda fit together at the classroom level, particularly those that related to the variables of teacher efficacy, teacher behavior, and student learning. The specific purposes of this study were to inquire into the relationships between teacher efficacy and changes in teacher behavior and student learning espoused by reform proponents and to examine the impact of a curriculum innovation that addressed reform elements on student learning and teacher efficacy.

While a fair amount of prior research had shown that teacher efficacy correlated both with teacher behavior and student learning, this study was designed to focus on more sophisticated pedagogical behaviors associated with setting high learner expectations, using challenging materials, and emphasizing higher-order thinking, problem-solving, and metacognitive skills in the classroom. Rather than measuring student learning gains through traditional standardized achievement instruments, the study investigated gains in concept learning, critical thinking, and mastery of standards of learning in keeping with the reform emphasis on multiple, performance-based assessments and content standards.

Research findings on the construct of efficacy and its relationship to teacher behavior and student learning proved inconclusive, but findings related to the impact of a curriculum innovation in social studies on observable teaching practice were more encouraging. Although not examined within the context of an experimental design, both teachers themselves and

external observers reported an increase in specific pedagogical strategies during implementation of pilot curriculum lessons. While these perceived changes in teacher behavior were not deep enough to translate into student learning gains at this stage of implementation, they built a solid baseline that can continue to be tracked as the project moves forward into year two.

Even though the study raised more questions than it was able to answer, it reenforced the importance of a research platform that begins to investigate more rigorously how some of the tenets of the reform agenda actually play out in public school classroom settings. Perhaps its most salient contribution was to echo the findings of other literature on the complexity of effecting deep and sustained change in the teaching-learning process. Despite having incorporated key features that facilitate change into the project's design including:

- use of teacher volunteers who committed to 25 hours of instruction,
- use of grade level teams to support collaboration,
- adherence to less is more (focus on 10-12 pilot lessons),
- alignment of the innovation with state and local curriculum frameworks and learning standards,
- allowance for flexibility in the adaptation of lessons and time frames at the classroom and building levels,
- provision of support through various modes of technical assistance as requested in the classroom,
- administrative commitment, and
- staff development structures that integrated content with relevant pedagogy,

many teachers were not able to see the congruence between the curriculum innovation and their district's goals and objectives and/or had difficulty applying the pilot lessons in their classrooms.

Thus, the overall implementation during the pilot phase of the project was limited in both scope and depth.

Nevertheless, the first wave of the curriculum development process proceeded quite smoothly. Feedback from teachers to developers led to many revisions in the lesson plans after this study was executed. These revised and expanded units will become the basis for continued research on this potentially promising curriculum innovation in social studies in future years.



## **APPENDIX: INSTRUMENTS USED IN THE STUDY**

## ***Project Phoenix Questionnaire***

### ***Section 1: Demographic Information***

1. Last Name: \_\_\_\_\_ First Name: \_\_\_\_\_ MI \_\_\_\_\_
2. Date of Birth: \_\_\_\_\_ 3. Gender: ☐ Male ☐ Female
4. Highest Degree Obtained (check one)  
☐ B.A. or B.S. ☐ M.A. or M.S. or M.Ed.. ☐ Ed.S. ☐ Ed.D. or Ph.D
5. Number of graduate level courses beyond highest degree obtained: \_\_\_\_\_
6. Number of graduate level courses in gifted education: \_\_\_\_\_ in social studies education \_\_\_\_\_
7. Do you have an endorsement in gifted education ☐ Yes ☐ No ☐ Currently working on endorsement
8. Role: ☐ Classroom teacher ☐ Media specialist
9. Total number of years teaching experience: \_\_\_\_\_
10. School in which you currently teach: \_\_\_\_\_
11. Number of years you have taught in that school: \_\_\_\_\_
12. Number of years you have taught in Norfolk district: \_\_\_\_\_
13. When were you hired into the Norfolk district: \_\_\_\_\_ since 1993-94 school year  
☐ 1992-93 school year or before
14. Grade level in which you currently teach: \_\_\_\_\_
15. Number of years you have taught at that level: \_\_\_\_\_

#### **General Directions**

Please read the directions for completing the next two scales. Each scale has a different rating system because it has been adapted from different instruments. These scales require a forced choice answer so you will need to pick the answer that best approximates your opinion. **Please answer every question because the analysis is based on a composite score.**

Your individual results will be treated confidentially. Only group data will be reported.

### ***Section II. General Teaching Efficacy Scale***

A number of statements about organizations, people, and teaching are presented below. The purpose is to gather information regarding the actual attitudes of educators concerning these statements. There are no correct or incorrect answers. We are interested only in your frank opinions.

**INSTRUCTIONS:** Please indicate your personal opinion about each statement by circling the appropriate response at the right of each statement.

**KEY: 1 = strongly agree. 2= moderately agree, 3= agree slightly more than disagree.  
 4= disagree slightly more than agree. 5= moderately disagree. 6= strongly disagree**

- |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| 1. The amount a student can learn is primarily related to family background.  | 1 | 2 | 3 | 4 | 5 | 6 |
| 2. If students aren't disciplined at home, they aren't likely to accept any discipline.   | 1 | 2 | 3 | 4 | 5 | 6 |
| 3. When I really try, I can get through to most difficult students.   | 1 | 2 | 3 | 4 | 5 | 6 |
| 4. A teacher is very limited in what he/she can achieve because a student's home environment is a large influence on his/her achievement. | 1 | 2 | 3 | 4 | 5 | 6 |

- |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| 5. If parents would do more for their children, I could do more.  | 1 | 2 | 3 | 4 | 5 | 6 |
| 6. If a student did not remember information I gave in a previous lesson, I would know how to increase his/her retention in the next lesson.                      | 1 | 2 | 3 | 4 | 5 | 6 |
| 7. If a student in my class becomes disruptive and noisy, I feel assured that I know some techniques to redirect him/her quickly.                                 | 1 | 2 | 3 | 4 | 5 | 6 |
| 8. If one of my students couldn't do a class assignment, I would be able to accurately assess whether the assignment was at the correct level of difficulty.      | 1 | 2 | 3 | 4 | 5 | 6 |
| 9. If I really try hard, I can get through to even the most difficult or unmotivated student.   | 1 | 2 | 3 | 4 | 5 | 6 |
| 10. When it comes right down to it, a teacher really can't do much because most of a student's motivation and performance depends on his or her home environment. | 1 | 2 | 3 | 4 | 5 | 6 |

### *Section III. Social Studies Teaching Efficacy Scale*

Please indicate the degree to which you agree or disagree with each statement below by circling the appropriate letters to the right of each statement

SA = Strongly Agree A = Agree UN = Uncertain D = Disagree SD = Strongly Disagree

- |   |    |   |    |   |    |
|---|----|---|----|---|----|
| 1. When a student does better than usual in social studies it is often because the teacher exerted a little extra effort.                     | SA | A | UN | D | SD |
| 2. I am continually finding better ways to teach social studies.  | SA | A | UN | D | SD |
| 3. Even when I try very hard, I don't teach social studies as well as I do most subjects.   | SA | A | UN | D | SD |
| 4. When the social studies grades of students improve, it is most often due to their teacher having found a more effective teaching approach. | SA | A | UN | D | SD |
| 5. I know the steps necessary to teach social studies concepts effectively.   | SA | A | UN | D | SD |
| 6. I am not very effective in monitoring social studies learning.   | SA | A | UN | D | SD |
| 7. If students are underachieving in social studies it is most likely due to ineffective social studies teaching.                             | SA | A | UN | D | SD |
| 8. I generally teach social studies ineffectively.  | SA | A | UN | D | SD |
| 9. The inadequacy of a student's social studies background can be overcome by good teaching.  | SA | A | UN | D | SD |
| 10. The low social studies achievement of some students cannot generally be blamed on their teachers.   | SA | A | UN | D | SD |
| 11. When a low achieving child progresses in social studies, it is usually due to extra attention given by the teacher.                       | SA | A | UN | D | SD |
| 12. I understand social studies concepts well enough to be effective in teaching elementary or middle school social studies.                  | SA | A | UN | D | SD |

13. Increased effort in social studies teaching produces little change in some students' social studies achievement	SA	A	UN	D	SD
14. The teacher is generally responsible for the achievement of students in social studies.	SA	A	UN	D	SD
15. Students' achievement in social studies is directly related to their teacher's effectiveness in social studies teaching.	SA	A	UN	D	SD
16. If parents comment that their child is showing more interest in social studies at school, it is probably due to the performance of the child's teacher.	SA	A	UN	D	SD
17. I find it difficult to explain different cultural perspectives in social studies.	SA	A	UN	D	SD
18. I am typically able to answer students' social studies questions.	SA	A	UN	D	SD
19. I wonder if I have the necessary skills to teach social studies.	SA	A	UN	D	SD
20. Effectiveness in social studies teaching has little influence on the achievement of students with low motivation.	SA	A	UN	D	SD
21. Given a choice, I would not invite the principal to evaluate my social studies teaching.	SA	A	UN	D	SD
22. When a student has difficulty understanding a social studies concept, I am usually at a loss as to how to help the student understand it better.	SA	A	UN	D	SD
23. When teaching social studies, I usually welcome student questions.	SA	A	UN	D	SD
24. I don't know what to do to turn students on to social studies.	SA	A	UN	D	SD
25. Even teachers with good social studies teaching abilities cannot help some kids learn social studies.	SA	A	UN	D	SD

#### *Section IV. Project Phoenix Applications*

SA = Strongly Agree A = Agree UN = Uncertain D = Disagree SD = Strongly Disagree

26. I find it easy to teach reasoning skills to students through the discipline of social studies.	SA	A	UN	D	SD
27. I find it difficult to use social studies content to support interpersonal and group process skills in students.	SA	A	UN	D	SD
28. I find it difficult to promote depth of learning in social studies.	SA	A	UN	D	SD
29. I find it easy to help students make' interdisciplinary connections among history, politics, economics, and geography.	SA	A	UN	D	SD
30. I find it difficult to embed mathematics and technology into social studies learning.	SA	A	UN	D	SD
31. I find it easy to enhance students' research skills through social studies.	SA	A	UN	D	SD



# External Observer Form

Revision Date: 27 Jan 1999

Name of Observer: \_\_\_\_\_ Number of Students: \_\_\_\_\_  
 Date: \_\_\_\_\_ Grade: \_\_\_\_\_ Length of Observation (minutes): \_\_\_\_\_  
 Name of School: \_\_\_\_\_  
 Name of Teacher: \_\_\_\_\_  
 Course/Subject Observed: ☐ Social Studies ☐ Mathematics  
 Classroom Desk Arrangement: ☐ desks in rows and columns ☐ desks grouped  
☐ Other (specify): \_\_\_\_\_

Please outline exactly what you are observing in the classroom with respect to curriculum and instruction. Describe

the specific lesson,  
 the organization of the lesson,  
 the texts and/or materials used,  
 the methods used in communicating the lesson,  
 characteristics of the learning experience and environment,

or any other observations and impressions which became the basis for completing the attached checklist.  
 The categories on the checklist are as follows:

Curriculum Planning	Expectations for Learning	Accom. for Individ. Differences
Curric. Delivery Features	Gen. Teaching Strategies	Critical Thinking Strategies
Problem-solving Strategies	Metacognition	Classroom Extensions

## Teacher Interview Questions

1. Do you have a written plan for this lesson?
2. What were your instructional objectives during the previous lesson with this class? What will you be covering in the subsequent lesson?
3. Are there any aspects of the lesson which you want to clarify with me before I finalize this observation form?
4. Observer specified question: \_\_\_\_\_

Area	The teacher...	Yes	No	Comments
Curriculum Planning	1. had a written lesson plan linked to course objectives.			
	2. communicated the purpose/objectives of the lesson to students.			
	3. adhered to the basic framework of the lesson as originally intended.			
Expectations for Learners	4. was clear in giving directions, and discussing activities and assignments.			
	5. sets high expectations for student performance in the classroom.			
	6. provided clear and consistent feedback on student performance.			
Accommodation to Indiv. Differences	7. presented content which challenged students.			
	8. accommodated individual or subgroup differences through material selection or task assignments.			
	9. incorporated multicultural perspectives or knowledge, reflecting at least two cultures.			
	10. addressed at least 2 different modes of learning, e.g. visual, auditory, kinesthetic.			
	11. allowed students individually or in small groups to move through basic material more rapidly.			
Curric. Deliv. Features	12. emphasized depth in learning.			
	13. taught according to key concepts and ideas relevant to content area being addressed.			
	14. encouraged or indicated interdisciplinary connections.			
General Teaching Strategies	15. used flexible patterns of grouping to deliver the lesson.			
	16. used more than one instructional strategy to deliver the lesson.			
	17. provided activities in which students applied new learning.			
	18. provided the opportunity for the students to use technology.			
	19. kept all or most of the students on task.			
	20. used hands-on approaches including such things as journaling, manipulatives, experiments, etc.			
	21. used cooperative or collaborative learning strategies.			
	22. allowed students to discover central ideas on their own through structured activities and/or questions.			
	23. emphasized higher level thinking strategies/skills.			
Critical Thinking Strategies	Used activities or questions which enabled students:			
	24. to make judgements or evaluate situations, problems, or issues.			
	25. to compare and contrast.			
	26. to generalize from specific data to the abstract.			
	27. to synthesize or summarize information within or across the disciplines.			
	28. to debate points of view or develop arguments to support ideas.			

Area	The teacher...	Yes	No	Comments
Problem-solving Strategies	Used activities or questions which encouraged students:			
	29. to brainstorm ideas or alternatives.			
	30. to define problems (to go from a "mess" to a well-defined problem statement).			
	31. to select and implement solutions to problems.			
	32. to explore multiple interpretations.			
	33. allowed students to use alternative rather than single modes of expression for class/homework activities/products (e.g., charts, graphics, videos, journals, etc.)			
	34. allowed students to self-select topics for further investigation.			
Metacognition	35. modeled metacognitive strategies such as planning, monitoring, self-reflection or self-appraisal.			
	36. provided opportunities for students to think about their own thinking.			
	37. had students reflect on their own performance.			
Classroom Extensions	38. reinforced or expanded the lesson by assigning homework.			
	39. provided follow-up ideas of special projects for students to pursue.			
	40. identified people or materials which could be used to supplement student learning.			

Observer's Signature: \_\_\_\_\_



The College Of  
**WILLIAM & MARY**



## Teacher Self-Report Form

Revision Date: 27 Jan 1999

Name of Observer: \_\_\_\_\_ Number of Students: \_\_\_\_\_  
 Date: \_\_\_\_\_ Grade: \_\_\_\_\_ Length of Observation (minutes): \_\_\_\_\_  
 Name of School: \_\_\_\_\_  
 Name of Teacher: \_\_\_\_\_  
 Course/Subject Observed: ☐ Social Studies ☐ Mathematics  
 Classroom Desk Arrangement: ☐ desks in rows and columns ☐ desks grouped  
☐ Other (specify): \_\_\_\_\_

Please outline exactly what you are observing in the classroom with respect to curriculum and instruction.

Describe

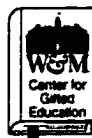
the specific lesson,  
 the organization of the lesson,  
 the texts and/or materials used,  
 the methods used in communicating the lesson,  
 characteristics of the learning experience and environment,

or any other observations and impressions which became the basis for completing the attached checklist.

The categories on the checklist are as follows:

Curriculum Planning	Expectations for Learning	Accom. for Indiv. Differences
Curric. Delivery Features	Gen. Teaching Strategies	Critical Thinking Strategies
Problem-solving Strategies	Metacognition	Classroom Extensions





Area	Behavior: I...	Yes	No	Comments
Curriculum Planning	1. had a written lesson plan linked to course objectives.			
	2. communicated the purpose/objectives of the lesson to students.			
	3. adhered to the basic framework of the lesson as originally intended.			
Expectations for Learners	4. was clear in giving directions, and discussing activities and assignmts.			
	5. sets high expectations for student performance in the classroom.			
	6. provided clear and consistent feedback on student performance.			
Accommodation to Indiv. Differences	7. presented content which challenged students.			
	8. accommodated individual or subgroup differences through material selection or task assignments.			
	9. incorporated multicultural perspectives or knowledge, reflecting at least two cultures.			
	10. addressed at least 2 different modes of learning, e.g. visual, auditory, kinesthetic.			
	11. allowed students individually or in small groups to move through basic material more rapidly.			
Curric. Deliv. Features	12. emphasized depth in learning.			
	13. taught according to key concepts and ideas relevant to content area being addressed.			
	14. encouraged or indicated interdisciplinary connections.			
General Teaching Strategies	15. used flexible patterns of grouping to deliver the lesson.			
	16. used more than one instructional strategy to deliver the lesson.			
	17. provided activities in which students applied new learning.			
	18. provided the opportunity for the students to use technology.			
	19. kept all or most of the students on task.			
	20. used hands-on approaches including such things as journaling, manipulatives, experiments, etc.			
	21. used cooperative or collaborative learning strategies.			
	22. allowed students to discover central ideas on their own through structured activities and/or questions.			
	23. emphasized higher level thinking strategies/skills.			
Critical Thinking Strategies	Used activities or questions which enabled students:			
	24. to make judgements or evaluate situations, problems, or issues.			
	25. to compare and contrast.			
	26. to generalize from specific data to the abstract.			
	27. to synthesize or summarize information within or across the disciplines.			
	28. to debate points of view or develop arguments to support ideas.			

Area	Behavior: I...	Yes	No	Comments
Problem-solving Strategies	Used activities or questions which encouraged students:			
	29. to brainstorm ideas or alternatives.			
	30. to define problems (to go from a "mess" to a well-defined problem statement).			
	31. to select and implement solutions to problems.			
	32. to explore multiple interpretations.			
	33. allowed students to use alternative rather than single modes of expression for class/homework activities/products (e.g., charts, graphics, videos, journals, etc.)			
	34. allowed students to self-select topics for further investigation.			
Metacognition	35. modeled metacognitive strategies such as planning, monitoring, self-reflection or self-appraisal.			
	36. provided opportunities for students to think about their own thinking.			
	37. had students reflect on their own performance.			
Classroom Extensions	38. reinforced or expanded the lesson by assigning homework.			
	39. provided follow-up ideas of special projects for students to pursue.			
	40. identified people or materials which could be used to supplement student learning.			

Teacher's Signature:

**The William & Mary Test  
Of  
Conceptual Thinking Abilities  
(CNTA)  
Grade 2**

**Developed by**

**Karen B. Rogers, Ph.D.  
University of St. Thomas**

**© 1999 by Karen B. Rogers**

NAME: \_\_\_\_\_

**Problem 1: How would you group these food items? Make THREE groups with at least TWO items in each group. No food can be put into more than one group. Name the three groups and list the items that belong to that group below the name.**

- Cheese
- Raisin Bran
- Oranges
- Oatmeal
- Chocolate milk
- Bananas
- Ice cream
- Peaches
- Corn Flakes

Group 1 \_\_\_\_\_ Items \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Group 2 \_\_\_\_\_ Items \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Group 3 \_\_\_\_\_ Items \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**Problem 2.** What are all the different ways you can rearrange these letters?

A B C

_____	_____	_____
_____	_____	_____

**Problem 3.** List the next four letters for each series.

SAMPLE PROBLEM: 1 2 3 4 5 6 ANSWER: 7 8 9 10

R S R S R S	_____	_____	_____	_____
M M M N N N	_____	_____	_____	_____
D E P F G P	_____	_____	_____	_____

**Problem 4.** Mr. Magician has 2 boxes, labeled “YES” and “NO.” He has cards with pictures that are:

- Diamonds or Triangles
- Purple or Orange
- Jagged or Squiggly
- 1 figure or 4 figures

He begins sorting his cards, putting each one in either the “YES” or “NO” box.

He does the first four cards. You do the rest.

The Card	YES	NO
1 purple jagged diamond		X
1 orange jagged diamond		X
4 purple squiggly diamonds	X	
4 purple jagged triangles		X
4 orange jagged diamonds		
1 purple jagged triangle		
4 orange squiggly triangles		
4 purple squiggly triangles		
4 orange jagged triangles		
4 orange squiggly diamonds		

**What is Mr. Magician’s Rule? (Fill in the Blanks.)**

Cards with \_\_\_\_\_ go in the YES box. The cards that are YES can also have \_\_\_\_\_, but they cannot have \_\_\_\_\_.

**Problem 5.** A **SYSTEM** is made up of many parts that are related to each other. Some of the parts make the other parts start going. These parts are called the **INPUT**. Other parts begin to get going once some of the parts begin to act on them. These parts are called the **PROCESS**. And when these parts are going, something happens. This is called the **OUTPUT**. This **OUTPUT** then leads back to the beginning **INPUT**.

A **Bicycle** is a good example of a **SYSTEM**. It has many parts, such as wheels, handlebars, drive chain, brakes, seat, and pedals. A bike rider sits on the seat, puts hands on the handlebars, and presses down on one pedal with one foot (**Input**). As the pedal goes down, the other pedal comes up and the drive chain rotates (**Process**), which makes the wheels go forward (**Output**). As the other pedal comes up, the rider presses down on it with the other foot and the process happens again and again.

**5A.** Which of these objects are also examples of a system? Circle Yes or NO.

Playing Tag	YES	NO
An envelope	YES	NO
A computer	YES	NO
Vacuum cleaner	YES	NO
A nail	YES	NO

**5B. Choose ONE object you identified as a SYSTEM from this list and explain in your own words how it is a system.**

---

---

---

---

---

---

---

---

---



**The William & Mary Test  
Of  
Conceptual Thinking Abilities  
  
(CNTA)  
  
Grade 4**

**Developed by**

**Karen B. Rogers, Ph.D.  
University of St. Thomas**

**© 1999 by Karen B. Rogers**

NAME: \_\_\_\_\_

**Problem 1: How would you group these food items? Make as many groups as you can but at least TWO items should belong to each group and you can use each item only once. Attach a name to each group and list the items that belong to that group below the name.**

- Cheese
- Oranges
- Peas
- Fruit juice bars
- Oatmeal
- Chocolate milk
- Lettuce
- Bananas
- Peaches
- Ice cream
- Corn Flakes
- Yogurt
- Apple juice
- Tomatoes
- Frozen pizza
- Raisin bran
- Bottled water
- Whipped cream

Group 1 \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Group 2 \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Group 3 \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Group 4 \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Group 5 \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Group 6 \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**Problem 2.** What are all the different ways you can rearrange these letters?

A B C D

_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

**Problem 3.** List the next four letters for each series.

SAMPLE PROBLEM: 1 2 3 4 5 6 ANSWER: 7 8 9 10

R S R S R S	_____	_____	_____	_____
M M M N N N	_____	_____	_____	_____
D E P F G P	_____	_____	_____	_____
T U F G U V G H	_____	_____	_____	_____
W X A X Y B	_____	_____	_____	_____
T D E T E F T F G	_____	_____	_____	_____

**Problem 4.** Mr. Magician has 2 boxes, labeled “YES” and “NO.” He has cards with pictures that are:

- Diamonds or Triangles
- Purple or Orange
- Jagged or Squiggly
- 1 figure or 4 figures

He begins sorting his cards, putting each one in either the “YES” or “NO” box.

He does the first four cards. You do the rest.

The Card	YES	NO
1 purple jagged diamond		X
1 orange jagged diamond		X
4 purple squiggly diamonds	X	
4 purple jagged triangles		X
4 orange jagged diamonds		
1 purple jagged triangle		
4 orange squiggly triangles		
4 purple squiggly triangles		
4 orange jagged triangles		
4 orange squiggly diamonds		
4 purple jagged diamonds		
1 orange squiggly triangle		

**What is Mr. Magician’s Rule? (Fill in the Blanks.)**

Cards with \_\_\_\_\_ go in the YES box. The cards that are YES can also have \_\_\_\_\_, but they cannot have \_\_\_\_\_.

**Problem 5.** A **SYSTEM** is made up of many parts that are related to each other. Some of the parts make the other parts start going. These parts are called the **INPUT**. Other parts begin to get going once some of the parts begin to act on them. These parts are called the **PROCESS**. And when these parts are going, something happens. This is called the **OUTPUT**. This **OUTPUT** then leads back to the beginning **INPUT**.

A Bicycle is a good example of a **SYSTEM**. It has many parts, such as wheels, handlebars, drive chain, brakes, seat, and pedals. A bike rider sits on the seat, puts hands on the handlebars, and presses down on one pedal with one foot (Input). As the pedal goes down, the other pedal comes up and the drive chain rotates (Process), which makes the wheels go forward (Output). As the other pedal comes up, the rider presses down on it with the other foot and the process happens again and again.

**5A.** Which of these objects are also examples of a system? Circle Yes or NO.

Playing Tag	YES	NO
A Computer	YES	NO
An envelope	YES	NO
Savings account	YES	NO
A classroom	YES	NO
A Nail	YES	NO
A fire engine	YES	NO

Conceptual Thinking  
Grade 4 Draft

**5B. Choose ONE object you identified as a SYSTEM from this list and explain in your own words how it is a system.**

---

---

---

---

---

---

---

---

---

**The William & Mary Test  
Of  
Critical Thinking Abilities  
  
(CRTA)  
  
Grade 2**

**Developed by**

**Karen B. Rogers, Ph.D.  
University of St. Thomas**

**© 1999 by Karen B. Rogers**

NAME: \_\_\_\_\_

**Problem 1A** Describe how each pair of words is related.**Example:** Fork      Spoon      Both are used to eat food.

Blanket      Bed      \_\_\_\_\_

Boat      Sail      \_\_\_\_\_

Gold      Silver      \_\_\_\_\_

Rose      Daisy      \_\_\_\_\_

Baby      Sleep      \_\_\_\_\_

Lid      Pan      \_\_\_\_\_

**Problem 1B.** What are different categories for at least two sets of the words above were related? Name two possible categories.

<b>Example:</b>	Fork	Spoon	<u>used for eating</u>
	Stove	Refrigerator	<u>appliance for food</u>
	<b>CATEGORY</b>		<u>Things found in the</u>
			<u>Kitchen</u>

1. \_\_\_\_\_

2. \_\_\_\_\_



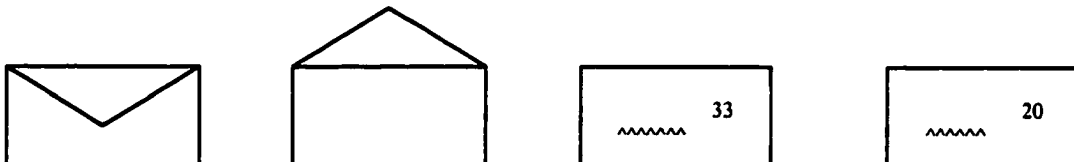
**Problem 2.** Some facts about islands are stated below. Circle the letters of **ONLY** those facts that **HELP TO ANSWER THE QUESTION**. There may be more than one letter to circle for the question.

**THE QUESTION:** Why are islands surrounded by water?

- A. Islands are the tips of huge mountains that are found on the ocean floor.
- B. Islands are green with lots of palm trees.
- C. When water pushes against rock, it gradually breaks the rock down into dirt and sand, forming an island.
- D. Islands can be dangerous for passing ships.
- E. People can live on islands.
- F. The definition of an island explains why it is surrounded by water.

**Problem 3.** Circle those envelopes you would need to turn over to **BE SURE** the rule was followed.

- **Rule:** A sealed letter must have a 33-cent stamp on it.



**Problem 4. Circle the word choice that matches the key word IN THE SAME WAY the first three words match.**

MATCHED WORDS		KEYWORD		WORD CHOICES
EXAMPLE: Root...trunk...branch		SHOE		DAISY TREE FOOT HAT is furthest way from the foot just like a branch is furthest away from the roots of a tree.
Small...smaller...smallest		Teenager		CAR  BABY  GRANDMA  DOG
Run...hurry...speed		CRY		TEARS  WHINE  FIGHT  CHILD
Stale..tales...slate		DIAL		LAI D  PHONE  CIRCLE  FINGER
Through...achoo...crew		PAID		BILL  MONEY  WADE  OFF

**Problem 5.** Think of all the good points and bad points for this situation. (The chart below has given you one example of each kind of point.)

**SITUATION:** The government has passed a law that all teddy bears made in the USA must be brown in color.

GOOD POINTS	BAD POINTS
<ul style="list-style-type: none"> <li>• Brown won't show the dirty if the bear gets dirty</li> </ul>	<ul style="list-style-type: none"> <li>• It is hard to tell your own bear from everyone else's.</li> </ul>

**The William & Mary Test  
Of  
Critical Thinking Abilities  
  
(CRTA)  
  
Grade 4**

**Developed by**

**Karen B. Rogers, Ph.D.  
University of St. Thomas**

**© 1999 by Karen B. Rogers**

NAME: \_\_\_\_\_

Problem 1A Describe how each pair of words is related.

Example:	Fork	Spoon	<u>Both are used to eat food.</u>
	Blanket	Bed	_____
	Boat	Sail	_____
	Hurry	Delay	_____
	Rose	Daisy	_____
	Baby	Sleep	_____
	Lid	Pan	_____
	Love	Hate	_____
	Lamp	Candle	_____
	Juice	Milk	_____
	Siren	Bell	_____

Problem 1B. What are different categories for at least two sets of the words above were related? Name two possible categories.

Example:	Fork	Spoon	<u>used for eating</u>
	Stove	Refrigerator	<u>appliance for food</u>
	CATEGORY		<u>Things found in the</u>
			<u>Kitchen</u>

1. \_\_\_\_\_

2. \_\_\_\_\_

**Problem 2.** For each of the two questions below, some facts are stated. Circle the letters of **ONLY** those facts that **HELP TO ANSWER THE QUESTION**. There may be more than one letter to circle for the question.

**QUESTION:** Why are islands surrounded by water?

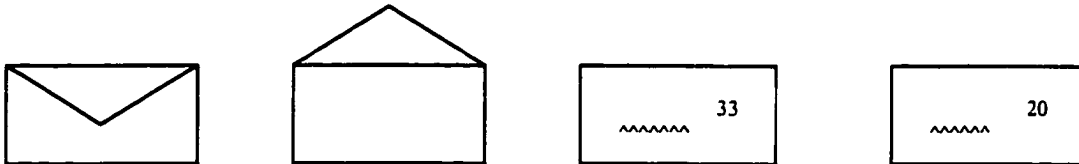
- A. Islands are the tips of huge mountains that are found on the ocean floor.
- B. Islands are green with lots of palm trees.
- C. When water pushes against rock, it gradually breaks the rock down into dirt and sand, forming an island.
- D. Islands can be dangerous for passing ships.
- E. People can live on islands.
- F. The definition of an island explains why it is surrounded by water.

**QUESTION:** Why does soap have to be used in a dishwasher?

- A. The moving water in the dishwasher gets rid of stock on food on the dishes.
- B. The combination of soap and moving water gets rid of greasy substances on the dishes.
- C. Most dirt cannot be removed only by water.
- D. The average dishwasher load contains 14 plates, 12 glasses, 3 pans, and 42 pieces of silverware.
- E. The heat or the water causes the soap to release cleaning and sterilizing agents.
- F. The average family uses 1 medium box of dishwasher soap per month.

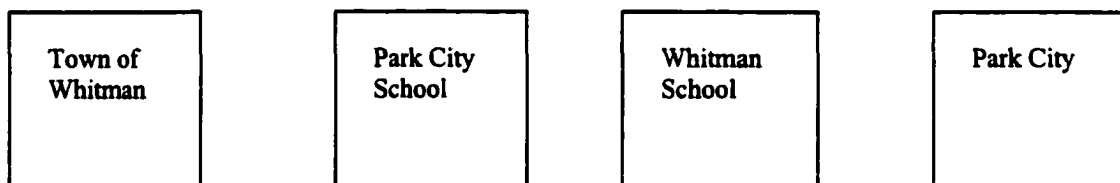
**Problem 3.** For each of the two problems below, circle those items you would need to turn over to **BE SURE** the rule was followed.

- **Rule:** A sealed letter must have a 33-cent stamp on it.



**The Rule:** A student who goes to Park City School must live in Park City.

Each student's school is placed on one side of the attendance card and the student's city of residence is on the other side of the card. Circle the cards you must turn over to be sure Park City students are the only ones going to Park City School.



**Problem 4. Circle the word choice that matches the key word IN THE SAME WAY the first three words match.**

MATCHED WORDS		KEYWORD		WORD CHOICES
EXAMPLE: Root...trunk...branch		SHOE		DAISY TREE FOOT HAT is furthest way from the foot just like a branch is furthest away from the roots of a tree.
Small...smaller...smallest		Teenager		CAR BABY GRANDMA DOG
Run...hurry...speed		CRY		TEARS WHINE FIGHT CHILD
Step...stair..staircase		LETTER		ENVELOPE STAMP ALPHABET SWEATER
Stale..tales...slate		DIAL		LAI PHONE CIRCLE FINGER
Through...achoo...crew		PAID		BILL MONEY WADE OFF



**Problem 5.** Think of all the good points and bad points for this situation. (The chart below has given you one example of each kind of point.)

**SITUATION:** The government has passed a law that all teddy bears made in the USA must be brown in color.

GOOD POINTS	BAD POINTS
<ul style="list-style-type: none"> <li>• Brown won't show the dirty if the bear gets dirty</li> </ul>	<ul style="list-style-type: none"> <li>• It is hard to tell your own bear from everyone else's.</li> </ul>

**The William & Mary Test  
Of  
Conceptual and Critical Thinking Abilities**

**Grade 7**

**Developed by**

**Karen B. Rogers, Ph.D.  
University of St. Thomas**

**© 1999 by Karen B. Rogers**

**Problem 2A**

Describe how each pair of words is related.

Blanket	bed	_____
Boat	sail	_____
Hurry	delay	_____
Rose	daisy	_____
Baby	sleep	_____
Lid	pan	_____
Love	hate	_____
Lamp	candle	_____
Juice	milk	_____
Siren	bell	_____
Sang	sing	_____
Tool	hammer	_____
Step	stair	_____
Animal	horse	_____
Went	goes	_____
Petal	flower	_____

**Problem 2B** What are the different categories for how things were related? Name at least three.

_____	_____	_____
_____	_____	_____

**Problem 3A.** Mr. Magician has two boxes. One says “YES” and one says “NO.” He also has cards with various figures on them. Some figures are diamonds and some are triangles, some figures are purple, and some figures are orange. Some figures are jagged and some are squiggly. Some cards have more than 1 figure on them and some only have 1 figure. Mr. Magician begins by showing you one card at a time and putting it in either the YES or NO box. After he sorts the first four cards for you, you are asked to sort the rest. Tell which box each of the remaining 12 cards should go into.

**Mr. Magician’s Four Cards:**

1 purple jagged diamond	NO
1 orange jagged diamond	NO
4 purple squiggly diamonds	YES
4 purple jagged triangles	NO

**Your cards:** YES NO

1 orange jagged triangle		
1 purple squiggly triangle		
1 orange squiggly triangle		
1 purple squiggly diamond		
4 orange jagged diamonds		
1 purple jagged triangle		
4 orange squiggly triangle		
4 purple squiggly triangles		
4 orange jagged triangles		
4 orange squiggly diamonds		
4 purple jagged diamonds		
1 orange squiggly triangle		

**Problem 3B.** What is the rule for sorting the cards into the YES box? Be sure to describe which characteristics of the card are important for being a yes and which are not important.

To be a YES, the card must be \_\_\_\_\_, but doesn’t need to be \_\_\_\_\_.

**Problem 4A.** A **SYSTEM** is made up of many parts that are related to each other. Some of the parts make the other parts start going. These parts are called the **INPUT**. Other parts begin to get going once some of the parts begin to act on them. These parts are called the **PROCESS**. And when these parts are going, something happens. This is called the **OUTPUT**.

A Bicycle is a good example of a **SYSTEM**. It has many parts, such as wheels, handlebars, drive chain, brakes, seat, and pedals. For input, a bike rider sits on the seat, puts hands on the handlebars, and presses down on one pedal with one foot (Input). As the pedal goes down, the other pedal comes up and the drive chain rotates (process), which makes the wheels go forward (output). As the other pedal comes up, the rider presses down on it with the other foot and the process happens again and again.

Which of these objects are also examples of a system? Circle Yes or No.

The alphabet	YES	NO
A book	YES	NO
A computer	YES	NO
Vacuum cleaner	YES	NO
A classroom	YES	NO
A mailing envelope	YES	NO
A fire engine	YES	NO
A t.v. commercial	YES	NO
A tree	YES	NO
Playing tag	YES	NO

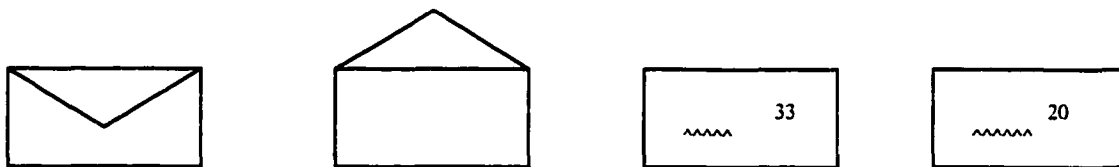
**4B. Choose ONE object you identified as a SYSTEM from this list and explain in your own words how it is a system.**

**Problem 5. For each of the two questions below, some facts are stated. Circle the letters of only those facts that help to answer the question. There may be more than one letter to circle for each question.**

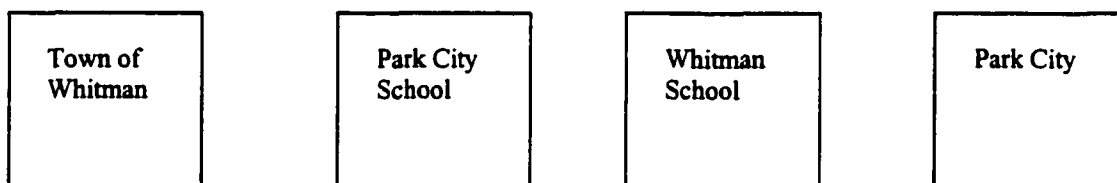
- **Why are islands surrounded by water?**
  - (a) Islands are the tips of huge mountains that are found on the ocean floor.
  - (b) Islands are green with lots of palm trees.
  - (c) When water pushes against rock, it gradually breaks the rock down into dirt and sand.
  - (d) Islands can be dangerous for passing ships.
  - (e) People can live on islands.
  - (f) The definition of an island explains why it is surrounded by water.
- **Why does dishwasher soap have to be used in a dishwasher?**
  - (a) The moving water in the dishwasher gets rid of stock on food on the dishes.
  - (b) The combination of soap and moving water gets rid of greasy substances on the dishes.
  - (c) Most dirt cannot be removed only by water.
  - (d) The average dishwasher load contains 14 plates, 12 glasses, 3 pans, and 42 pieces of silverware.
  - (e) The heat or the water causes the soap to release cleaning and sterilizing agents.
  - (f) The average family uses 1 medium box of dishwasher soap per month.

**Problem 6.** For each of the three problems below, circle those items you would need to turn over to **BE SURE** the rule was followed.

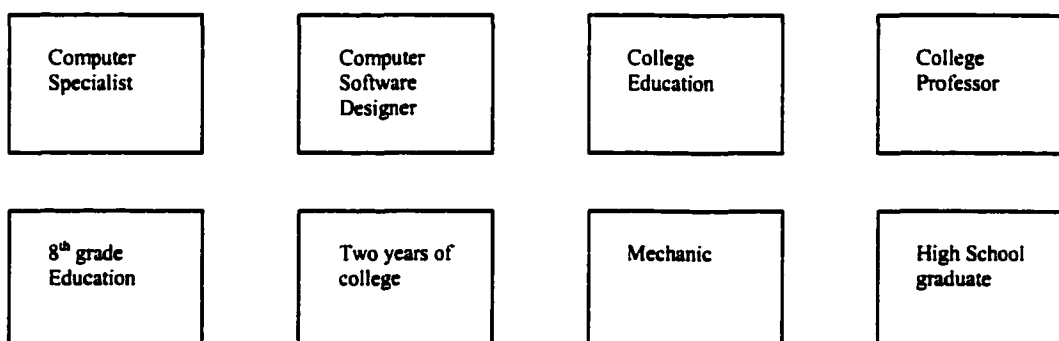
- **Rule:** A sealed letter must have a 33-cent stamp on it.



- **Rule:** A student who goes to Park City School must live in Park City. A student's school is placed on one side of the attendance card and the student's city of residence is on the other side of the card. Circle the cards you must turn over to be sure Park City students are the only ones going to Park City School.



- **Rule:** Only college graduates can work with computers. The career counselor has cards on all her clients, with the client's education level on one side of the card and the career in which the client was placed on the other. Circle the cards she must turn over to be sure the rule was not broken.



**Problem 7.** Which of the four word choices matches the keyword in the same way the first three words match each other? **CIRCLE YOUR CHOICE.**

MATCH WORDS	KEYWORD	CHOICES			
Small smaller smallest	Teenager	Car	Baby	Grandma	Dog
Run hurry speed	Cry	Tears	whine	fight	child
Root trunk branch	Shoe	Coat	tree	foot	hat
Stale tales slate	Dial	Laid	phone	circle	finger
Through achoo crew	Paid	Bill	money	wade	off



**Problem 8.** Think of all the good points, bad points, and interesting points for this situation (The chart below has given you one example of each kind of point).

**SITUATION:** The government has passed a law that all teddy bears made in the USA must be brown in color.

GOOD POINTS	BAD POINTS	INTERESTING POINTS
<ul style="list-style-type: none"> <li>Brown won't show the dirt if the bear gets dirty.</li> </ul>	<ul style="list-style-type: none"> <li>It is hard to tell your own bear from everyone else's.</li> </ul>	<ul style="list-style-type: none"> <li>We would see lots of different shades of brown in bears.</li> </ul>

### References

- Ahrne, G. (1994). Social organizations. Thousand Oaks, CA: Sage.
- Alexander, P. A., Murphy, P. K., & Woods, B. S. (1996). Of squalls and fathoms: Navigating the seas of educational innovation. Educational Researcher, 25(3), 31-36, 39.
- Anderson, R., Greene, M., & Loewen, P. (1988). Relationships among teachers' and students' thinking skills, sense of efficacy, and student achievement. Alberta Journal of Educational Research, 34(2), 148-165.
- Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., & Zellman, G. (1976). Analysis of the school preferred reading programs in selected Los Angeles minority schools (Report No. R-2007-LAUDS). Santa Monica, CA: RAND. (ERIC Document Reproduction Service No. 130 243)
- Ashton, P. (1985). Motivation and the teacher's sense of efficacy. In C. Ames & R. Ames (Eds.), Research on motivation in education, Vol. 2 (pp.141-171). Orlando, FL: Academic Press.
- Ashton, P. T., & Webb, R. B. (1986). Making a difference: Teachers sense of efficacy and student achievement. New York: Longman.
- Baldwin, A. Y. (1985, April). Identification and programming for black gifted children. Paper presented at the annual convention of the Council for Exceptional Children, Anaheim, CA. (ERIC Document Reproduction Service No. 261 495)
- Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is -or might be- the role of curriculum materials in teacher learning and instructional reform. Educational Researcher, 25 (9), 6-8, 14.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change.

Psychological Review, 84, 191-215.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: W. H. Freeman.

Benbow, C. P. (1998). Grouping intellectually advanced students for instruction. In J. VanTassel-Baska (Ed.), Excellence in educating gifted and talented learners (3rd ed., pp. 261-278). Denver, CO: Love Publishing.

Berk, R. A. (1998). Review of the Stanford achievement test (9<sup>th</sup> Ed). In J. C. Impara & B. S. Plake (Eds.), The 13<sup>th</sup> mental measurements yearbook (pp. 925-928). Lincoln, NE: University of Nebraska Press.

Berman, P., McLaughlin, M., Bass, G., Pauly, E., & Zellman, G. (1977). Federal programs supporting educational change: Vol. VII. Factors affecting implementation and continuation (Report No. R-1589/7-HEW). Santa Monica, CA: RAND. (ERIC Document Reproduction Service No. 140 432)

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. Teaching and Teacher Education, 13(3), 259-278.

Borko, H., & Putnam, R. T. (1996). Learning to teach. In D. C. Berliner & R. C. Calfee (Eds.), Handbook of Educational Psychology (pp.673-708). New York: Simon & Schuster Macmillan.

Callahan, C. M. (1996). A critical self-study of gifted education: Healthy practice, necessary evil, or sedition? Journal for the Education of the Gifted, 19, 148-163.

Cawalti, G. (1995). High school restructuring: What are the critical elements? NASSP Bulletin, 79 (569), 1-15.

Clasen, D. R. (1992). Changing peer stereotypes of high-achieving adolescents. NASSP Bulletin, 76(543), 95-102.

Cohen, D. K., & Spillane, J. P. (1993). Policy and practice: The relationship between governance and instruction. In S. H. Fuhrman (Ed.), Designing coherent education policy (pp.35-95). San Francisco: Jossey-Bass.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2<sup>nd</sup> ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cooper, H. M., Burger, J. M., & Seymour, G. E. (1979). Classroom context and student ability as influences on teachers' perceptions of classroom control. American Educational Research Journal, 16, 198-196.

Culross, R. R. (1997). Concepts of inclusion in gifted education. Teaching Exceptional Children, 29(3), 24-26.

Curda, L. K. (1997). Validation of a measure of teachers' efficacy and outcome expectations in the content domains of reading and mathematics. Unpublished doctoral dissertation. The University of Oklahoma, Tulsa.

Delisle, J. (1994). The inclusion movement is here-good . . it's about time. Au contraire. Gifted Child Today, 17(4), 30-31.

Epstein, J. (1995). School, family, community partnerships. Phi Delta Kappan, 76, 701-712.

Ford, D.Y. (1996). Reversing underachievement among gifted black students: Promising programs and Practices. New York: Teachers College Press.

Ford-Harris, D. Y., Schuerger, J. M., & Harris, J. J. (1991). Meeting the psychological needs of gifted Black students: A cultural perspective. Journal of counseling and Development, 69, 577-579.

Frasier, M. M. (1993). Issues, problems and programs in nurturing the disadvantaged and culturally different talented. In K. A. Heller, F. J. Monks, & A. H. Passow (Eds.), Research and development of giftedness and talent (pp. 685-692). New York: Pergamon.

Fullan, M. G. (1991). The new meaning of educational change. Toronto, Canada: Ontario Institute for Studies in Education.

Fullan, M. (1993). Change forces: Probing the depths of educational reform. New York: The Falmer Press.

Gagne, F. (1993). Constructs and models pertaining to exceptional human abilities. In K. A. Heller, F. J. Monks, & A. H. Passow (Eds.), International handbook of research and development of giftedness and talent (pp. 69-87). New York: Pergamon.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Educational research: An introduction (6th ed.). White Plains, NY: Longman Publishers.

Gardner, H. (1991). The unschooled mind. New York: Basic Books.

Gersten, R., & Guskey, T. R. (1985). Transforming teacher reluctance to teacher commitment. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service 260 068)

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. Journal of Educational Psychology, 76, 569-582.

Grimm, L. G., & Yarnold, P. R. (1995). Introduction to multivariate statistics. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics. Washington,

DC: American Psychological Association.

Guskey, T. R. (1981). The relationship of affect toward teaching and teaching self-concept to responsibility for student achievement. Journal of Social Studies Research 5 (2), 69-74.

Guskey, T. R. (1982). The effects of change in instructional effectiveness on the relationship of teacher expectations and student achievement. Journal of Educational Research, 75, 345-349.

Guskey, T. R. (1984). The influence of change in instructional effectiveness upon the affective characteristics of teachers. American Educational Research Journal, 21, 245-259.

Guskey, T. R. (1986). Staff development and process of teacher change. Educational Researcher, 15(5), 5-12.

Guskey, T. R. (1987). Context variables that affect measures of teacher efficacy. Journal of Educational Research, 81, 41-47.

Guskey, T. R. (1990). Integrating innovations. Educational Leadership, 47(5), 11-15.

Guskey, T. R. (1994). Results-oriented professional development: In search of an optimal mix of effective practices. Journal of Staff Development, 15(4), 42-50.

Guskey, T. R., & Passaro, P. D. (1994). Teacher efficacy: A study of construct dimensions. American Educational Research Journal, 31, 627-641.

Guskey, T. R., & Peterson, K. D. (1996). The road to classroom change. Educational Leadership, 53(4), 10-14.

Guskey, T. R., & Sparks, D. (1996). Exploring the relationship between staff development and improvements in student learning. Journal of Staff Development, 17(4), 34-38.

Havelock, R. G. (1969). A comparative study of the literature on the dissemination and utilization of scientific knowledge. Ann Arbor, MI: Center for Research on Utilization of Scientific Knowledge. (ERIC Document Reproduction Service No. 029 171)

Havelock, R. (1977). Solving educational problems: The theory and reality of innovation in developing countries. New York: Murray Hill Station.

Hoover-Dempsey, K. V., & Sandler, H. M. (1995). Parental involvement in children's education: Why does it make a difference? Teachers College Record, 97(2), 227-239.

Hoy, W. K., & Woolfolk, A. E. (1993). Teachers sense of efficacy and the organizational health of schools. The Elementary School Journal, 93, 356-372.

Joyce, B., & Showers, B. (1995). Student achievement through staff development (2nd ed.). White Plains, NY: Longman Publishing.

Klein, S. S. (1993). Are there better ways to identify and share the best within and among federal education program? Evaluation and Program Planning, 16, 227-239.

Kulik, J. A., & Kulik, C. C. (1992). Meta-analytic findings on grouping programs. Gifted Child Quarterly, 36, 73-77.

Margolin, L. (1994). Goodness personified: The emergence of gifted children. New York: Aldine De Gruyter.

Maddux, J. E. (1995). Self-efficacy theory: An introduction. In J. E. Maddux (Ed.), Self-efficacy, adaptation, and adjustment: Theory, research, and application (pp. 3-33). New York: Plenum Press.

Marzano, R. J. (1992). Cultivating thinking in English. Urbana, IL: National Council of Teachers of English.

Maker, C. J. (1996). Identification of gifted minority students: A national problem, needed changes and a promising solution. Gifted Child Quarterly, 40, 41-50.

Maker, C. J., & Schiever, S. W. (Eds.). (1989). Critical issues in gifted education: Defensible programs for cultural and ethnic minorities. Austin, TX: Pro-Ed.

McLaughlin, M. W., & Shepard, L. A. (1995). Improving education through standards-based reform. Stanford, CA: The National Academy of Education.

Moore, W., & Esselman, M. (1992, April). Teacher efficacy, power, school climate, and achievement: A desegregating district's experience. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Moye, V. H. (1997). Conditions that support transfer for change. Arlington Hts, IL: Skylight.

Myrtle, R. (1997, March). Implementing the William and Mary language arts units: In a school and across a school system. Presentation given at the Second Annual National Curriculum Network Conference, Williamsburg, VA.

O'Day, J. A., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. H. Fuhrman (Ed.), Designing coherent education policy (pp.250-312). San Francisco: Jossey-Bass.

Ogbu, J. U. (1994). Understanding cultural diversity and learning. Journal for the Education of the Gifted, 16, 355-383.

Olszewski-Kubilius, P., Grant, B., & Seibert, C. (1994). Social support systems and the disadvantaged gifted: A framework for developing programs and services. Roper Review, 17, 20-25.



Pajares, F. (1996). Self-efficacy beliefs in academic settings. Review of Educational Research, 66, 543-578.

Passow, A. H. (1990). Guest editorial. Roeper Review, 12, 134-136.

Passow, A. H. (1991). Educational programs for minority/disadvantaged gifted students. In R. Jenkins-Friedman, E. S. Richert, & J. F. Feldhusen (Eds.), Special populations of gifted learners: A book of readings (pp. 1-19). Washington, DC: National Association for Gifted Children.

Patton, J. M., & Baytops, J. L. (1995). Identifying and transforming the potential of young, gifted African-Americans: A clarion call for action. In B. A. Ford, F. E. Obiakor, & J. M. Patton (Eds.), Effective education of African-American exceptional learners (pp. 27-66). Austin, TX: Pro-Ed.

Paul, R. (1992). Critical thinking: What every person needs to survive in a rapidly changing world. Sonoma, CA: Critical Thinking Foundation

Payne, R. S. (1994). The relationship between teachers' beliefs and sense of efficacy and their significance to urban LSES minority students. Journal of Negro Education, 63(2), 181-196.

Perkins, D., & Saloman, G. (1989). Are cognitive skills context-bound? Educational Researcher, 18(1), 16-25.

Phillips, M. (1997). What makes schools effective? A comparison of the relationships of communitarian climate and academic climate to mathematics achievement and attendance during middle school. American Educational Research Journal, 34, 633-662.

Pogrow, S. (1998a). What is an exemplary program and why should anyone care? Educational Researcher, 27(7), 22-29.

Pogrow, S. (1998b). Systematically using learning environments to accelerate the learning of educationally disadvantaged students in grades 4-8. In C. Reigluth (Ed.), Instructional design theories and models: A new paradigm of instruction (in press). Mahwah, NJ: Lawrence Erlbaum Publishers.

Renzulli, J. S. (1977). The enrichment triad model: A guide for developing defensible programs for the gifted and talented. Wethersfield, CT: Creative Learning Press.

Renzulli, J. S. (1996). Schools for talent development: A practical plan for total school improvement. The School Administrator, 53, 20-22.

Riggs, I. (1995, April). The characteristics of high and low efficacy teachers. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco, CA.

Riggs, I., & Enochs, L. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. Science Education, 74, 625-638.

Rogers, K. B. (1999). Test analyses for Project Phoenix. Unpublished manuscript, College of William and Mary, Williamsburg, VA.

Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcomes. Journal of Educational Research, 74, 185-190.

Ross, J. A. (1992). Teacher efficacy and the effect of coaching on student achievement. Canadian Journal of Education, 17, 51-65.

Ross, J. A. (1995). Strategies for enhancing teachers' beliefs in their effectiveness: Research on a school improvement hypothesis. Teachers College Record, 97, 227-251.

Ross, J. A. (1998). Antecedents and consequences of teacher efficacy. In J. Brophy (Ed.), Advances in Research on Teaching (Vol. 7, pp. 49-74). Greenwich, CT: JAI Press.

Rubeck, M. L., & Enochs, L. G. (1991, April). A path analytic model of variables that influence science and chemistry teaching self-efficacy and outcome expectancy in middle school science teachers. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Lake Geneva, WI.

Sapon-Shevin, M. (1995). Why gifted students belong in inclusive schools. Educational Leadership, 52(4), 64-68.

Schunk, D. M. (1995). Self-efficacy and education and instruction. In J. E. Maddux (Ed.), Self-efficacy, adaptation, and adjustment: Theory, research, and application (pp. 281-303). New York: Plenum Press.

Senge, P. M. (1990). The fifth discipline: The art and practice of the learning organization. New York: Doubleday Currency.

Sergiovanni, T. J., & Starratt, R. J. (1993). Supervision: A redefinition (5<sup>th</sup> ed). New York: McGraw-Hill.

Shields, P. M., & Knapp, M. S. (1997). The promise and limits of school based reform: A national snapshot. Phi Delta Kappan, 79(4), 288-294.

Showers, B., Joyce, B., & Bennett, B. (1987). Synthesis of research on staff development: A framework for future study and a state-of-the-art analysis. Educational Researcher, 45, 77-87.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. Harvard Educational Review, 57, 1-22.

Simon, M. A., & Schifter, D. (1993). Toward a constructivist perspective: The impact of a mathematics teacher inservice program on students. Educational Studies in Mathematics, 25, 331-340.

Slavin, R. E. (1997). Design competitions: A proposal for a new federal role in educational research and development. Educational Researcher, 26(1), 22-29.

Sparks, D. (1994). A paradigm shift in staff development. Journal of Staff Development, 15, 26-29.

Sparks, D. (1995). Focusing staff development on improving student achievement. In G. Cawalti (Ed.), Handbook of research on improving student achievement (pp. 163-169). Arlington, VA: Educational Research Service.

Stanley, J. C. (1976). Identifying and nurturing the intellectually gifted. Phi Delta Kappan, 58(3), 234-238.

Stein, M. K., & Wang, M. C. (1988). Teacher development and school improvement: The process of teacher change. Teaching and Teacher Education, 4, 171-187.

Taba, H. (1962). Curriculum development: Theory and practice. New York: Harcourt, Brace, & World.

Tomlinson, C. A., & Callahan, C. M. (1992). Contributions of gifted education to general education in a time of change. Gifted Child Quarterly, 36, 183-189.

Tomlinson, C. A., Callahan, C. M., & Lelli, K. M. (1997). Challenging expectations: Case studies of high potential, culturally diverse young children. Gifted Child Quarterly, 41, 5-17.

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. Review of Educational Research, 68(2), 202-248.

United States Department of Education, National Commission on Excellence in Education. (1983). A nation at risk: An imperative for educational reform. Washington, DC: Author.

United States Department of Education, Office of Educational Research and Improvement. (1990). Accelerating academic achievement: A summary of findings from 20 years of NAEP. Washington, DC: Author.

United States Department of Education, Office of Educational Research and Improvement. (1993). National excellence: A case for developing America's talent. Washington, DC: Author.

VanTassel-Baska, J. (1992a). Planning effective curriculum for gifted learners. Denver, CO: Love Publishing.

VanTassel-Baska, J. (1992b). Educational decision making in acceleration and grouping. Gifted Child Quarterly, 36(2), 68-72.

VanTassel-Baska, J. (1995). The development of talent through curriculum. Roeper Review, 18, 98-102.

VanTassel-Baska, J., Avery, L. D., & Hall, D. (1997). Focus on the future: An evaluation of Ohio gifted programs, Vol. I. Unpublished report, College of William and Mary, Williamsburg, VA.

VanTassel-Baska, J., Bass, G. M., Ries, R. R., Poland, D. L., & Avery, L. D. (1998). A national pilot study of science curriculum effectiveness for high ability students. Gifted Child Quarterly, 42, 200-211.

VanTassel-Baska, J., & Brown, E. F. (in press). An analysis of gifted education curriculum models. In F. A. Karnes, & S. M. Bean (Eds.), Methods and materials for teaching the gifted and talented. Waco, TX: Prufrock Press.

VanTassel-Baska, J., Feldhusen, J., Seeley, K., Wheatley, G., Silverman, L., & Foster, W. (1988). Introduction. In J. VanTassel-Baska et al. (Eds.), Comprehensive curriculum for gifted

learners. Boston: Allyn and Bacon.

VanTassel-Baska, J., Hammett-Hall, K., & Bailey, J. M. (1996). Case studies of promising change schools. Research in Middle Level Education Quarterly, 19(2), 89-116.

VanTassel-Baska, J., Johnson, D. T., Hughes, C. E., & Boyce, L. N. (1996). A study of the language arts curriculum effectiveness with gifted learners. Journal for the Education of the Gifted, 19, 461-480.

VanTassel-Baska, J., Patton, J., & Prillaman, D. (1991). Gifted youth at risk: A report of a national study. Reston, VA: The Council for Exceptional Children.

Vinovskis, M. A. (1993). Analysis of the quality of research and development at the OERI research and development centers and at the OERI regional education laboratories. Washington, DC: U.S. Department of Education.

Ward, Jr., T. J., Ward, S. C., Landrum, M., & Patton, J. (1992, April). Examination of a new protocol for the identification of at-risk gifted learners. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. 387 979)

Watson, S. H. (1991). A study of the effects of teacher efficacy on the academic achievement of third-grade students in selected elementary schools in South Carolina. Unpublished doctoral dissertation, South Carolina State University, Orangeburg.

Werner, E. E., & Smith, R. S. (1992). Overcoming the odds: High risk children from birth to adulthood. Ithaca, NY: Cornell University.

Westberg, K. L., Archambault, F. X., Dobyns, S. M., & Salvan, T. J. (1993). The classroom practices observation study. Journal for the Education of the Gifted, 16, 120-146.

Wirt, F. M., & Kirst, M. W. (1997). The political dynamics of American education. Berkeley, CA: McCutchan Publishing Corporation.

Woolfolk, A. E., & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. Journal of Educational Psychology, 82(1), 81-91.

Wong, B. Y. L. (1997). Clearing hurdles in teacher adoption and sustained use of research-based instruction. Journal of Learning Disabilities, 30, 482-485.

## Vita

Linda Diane Avery

**Birthdate:** May 1, 1947

**Birthplace:** Quincy, Massachusetts

**Education:**

1996-1999	The College of William and Mary Williamsburg, Virginia Ph.D. in Educational Policy, Planning, and Leadership
1974-1975	Michigan State University East Lansing, Michigan M.A. in Educational Psychology
1965-1969	Baldwin-Wallace College Berea, Ohio B.A. in English Literature